



3-D Surface Reconstruction Using Spatial Frequency-based Approaches Under Influence of Perspective Distortion

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur

(Dr.-Ing.)

von M.Sc. Sherif Said Aly El-Etriby

geb. am 10. August 1970 in Tanta, Ägypten

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik
der Otto-von-Guericke-Universität Magdeburg

Gutachter:

Prof. Dr.-Ing. habil. Bernd Michaelis

Prof. Dr. rer. nat. habil. Peter Hauptmann

Promotionskolloquium am: 14.03.2008

This work is dedicated to ...
my parents, my wife (Saheir) and my children (Rawan, Rana and Ahmad)

Sherif

Abstract

The use of local spatial frequency provides a powerful analytical tool for image analysis. This dissertation provides an improved solution to long-standing problems in stereo vision; foreshortening, ambiguous matches, detecting and handling discontinuities and occlusion, and quantitative evaluation of stereo results.

Challenges arise from the fact that stereo images are acquired from slightly different views. Therefore, the projection of the surface in the image plane is more compressed and occupied a smaller area in one view than the other. This effect makes the matching of its two images very difficult and leads to confusing results. That is because, while corresponding two images N pixels on a scanline in one image may correspond to a different number of M pixels in the other image.

In this research, a new approach called *local-spatial-frequency* approach is proposed to combine the localizability of the spatial approach and the analytical benefits of the frequency approach. To simplify the matching process the prescribed system consists of a combination of stereo vision concept and the structured light concept. We also provide a solution for the long-standing problems in stereovision in two suggested algorithms:

The first algorithm is based on the output of linear spatial filters tuned to a range of orientations and scales that make the correspondence analysis more reliable and robust. The responses of these filters at a given pixel constitute a vector called filter response vector (FRV). This vector is correlated instead of correlating area in the two images. The correspondence problem can be solved by seeking points in the other view where this vector is maximally similar. In addition, an automatic procedure is used to evaluate and optimize the filters set by using the Steering theorem and the singular value decomposition (SVD). The projective distortion regions are detected to improve the quality of the disparity estimation by adapting to the size of filter kernel.

One of the major contributions of this algorithm appears while detecting and handling the depth discontinuities in order to improve the quality of the initial estimate disparity map. The algorithm maintains a current best estimate of the viewing parameters (to constrain vertical disparity to be consistent with epipolar geometry), a visibility map (to record whether a point is binocularly visible or occluded) and a

scale map (to record the largest scale of filter not straddling a depth discontinuity).

Starting with an initial computed disparity map, the algorithm iteratively updates the disparity for each detected region by adapting the size of filter kernel. The 3-D surface reconstruction can be calculated by the standard triangulation method. The experimental result shows that remarkable improvement is obtained in the projective distortion region.

The second suggested algorithm is a phase-difference based algorithm that uses an adaptive scale selection process. This algorithm demonstrates a theory of modeling the physical effects of perspective distortion (*foreshortening* problem) in stereo vision system. The central part of our model is the development of the dual scale factor that allow the reasoning of foreshortening in both the geometric domain of the world model and the frequency domain of the stereo images. The algorithm also provides a novel solution to the phase-wraparound problem that has limited the applicability of other phase-based methods.

This algorithm combines the magnitude and phase information for estimating depth information from two-dimensional stereo image pairs. This method takes into account not only the instability of phase but also the surface perspective distortion (the foreshortening in one view). These properties are important to the use of phase information to avoid the incorrect disparity estimates. Instead of matching intensities directly, a Gabor scale-space expansion is used. Magnitude information is used to detect "weak points" in the frequency domain, and only reliable phase values remain for a robust estimation disparity. The advantage of this algorithm is that the computed disparity values are obtained with sub-pixel accuracy without requiring explicit sub-pixel signal reconstruction. This relates the parameters in the image plane to the surface slope and does not require prior knowledge of the distance to the object. From the experimental results we conclude the fact that the foreshortening factor has its greatest impact when objects are sharply slanted and located near the cameras.

The efficiency and performance is confirmed on the basis of analysis of rectified stereo images. The experimental results show that the performance of the proposed algorithm in terms of accuracy and density of the disparity estimates has greatly improved. The random error could be determined by measuring reference (ground truth), for instance in the experimental results of cylinder object the error amounts to approximately ± 0.2 pixels while in the Area-based algorithm is approximately ± 1.2 pixels. Also for a center scanline from the slanted flat surface at the various rotation angles; 0° , 20° and 45° and disparity range (1...6) pixels, the error amounts are approximately ± 2.5 pixels.

Zusammenfassung

Für einfache Situationen gibt es viele leistungsfähige Verfahren zur 3D-Vermessung im Orts-/Zeit und Frequenzbereich, die in intensitätsbasierte wie auch merkmalsbasierte Verfahren unterteilt werden können. Diese versagen häufig unter dem Einfluss von nicht-kooperativen Messsituationen und erfüllen somit die Forderungen nach Robustheit und Flexibilität nur zum Teil. Die Transformation in den Ortsfrequenzbereich liefert ein leistungsfähiges Werkzeug für Bildanalyse, das in der vorgelegten Arbeit umfassend genutzt wird. Insbesondere bei komplizierten Objekten ergeben sich Probleme aus Tatsache, dass die Stereo-Bilder aus etwas unterschiedlichen Ansichten akquiriert werden. Deshalb ist die Projektion der Objektoberfläche in die Bilder aus einer Ansicht gestaucht und überdeckt ein kleineres Areal als aus der anderen Ansicht. Dieser Effekt erschwert das Matching beider Bilder und erfordert oft zusätzliche Massnahmen. Die Forschungsarbeiten im Rahmen dieser Dissertation liefern signifikant verbesserte Lösungsansätze für schon lange bestehende Korrespondenzprobleme bei der Stereobildanalyse: ortsabhängige Verzerrung in beiden Bildern, mehrdeutiges Matching, Behandlung von Unstetigkeiten und Verdeckungen sowie eine eindeutige Bewertung der Ergebnisse.

In dieser Arbeit wird eine hier als lokaler Ortsfrequenzansatz bezeichnete Herausgehensweise vorgeschlagen, die die Lokalisierbarkeit des räumlichen Ansatzes mit den analytischen Vorteilen des Frequenzansatzes kombiniert. Bei Objekten mit unzureichender Eigentextur und zur Optimierung der Messung wird durch Projektion strukturierten Lichts eine künstliche Textur generiert. In der vorgelegten Arbeit wird eine verbesserte Lösung zu oben genannten Problemen der Stereo-Vision durch Anwendung folgenden Algorithmen vorgeschlagen:

Der erste Algorithmus basiert auf der Anwendung linearer Raumfilter, die auf einen Bereich von Orientierungen und Skalierungen abgestimmt wurden, um die Korrespondenzanalyse zuverlässiger und robuster zu machen. Die Ausgangsinformationen dieser Filter an einem gegebenen Pixel bilden einen Vektor, der als Filterantwortvektor (FRV, filter response vector) bezeichnet wird. Dieser Vektor wird an Stelle einer Flächenkorrelation zwischen zwei Bildauschnitten korreliert. Das Korrespondenzproblem kann gelöst werden, indem der Punkt in der anderen Ansicht gesucht werden, an dem die Ähnlichkeitsfunktion des Vektors einen maximalen Wert

liefert. Zusätzlich wird eine automatische Prozedur verwendet, den Filtersatz auszuwerten und zu optimieren. Dies erfolgt unter Anwendung des Steering-Theorems und der Singulärwertzerlegung (SVD, singular value decomposition). Die projektiv verzeichneten Regionen werden erkannt und durch eine adaptive Grössenanpassung des Filterkerns kann die Qualität der Disparitätsschätzung verbessert werden. Einer der Hauptvorteile dieses Algorithmus ist das Finden und Behandeln der Tiefendiskontinuitäten zur qualitativen Verbesserung der initial geschätzten Disparitätskarte. Angefangen mit einer initialen Schätzung der Disparitätskarte aktualisiert der Algorithmus iterativ die Disparität für jede erkannte Region durch Anpassung der Grösse des Filterkerns. Die experimentellen Ergebnisse zeigen, dass eine signifikante Verbesserung in der projektiv verzeichneten Region erreicht wird.

Der zweite vorgeschlagene Algorithmus ist ein auf der Phasendifferenz basierender Algorithmus, der einen adaptiven Skalierungsauswahlprozess verwendet. Der Algorithmus nutzt eine Methode zur Modellierung der physikalischen Effekte perspektivischer Verzeichnungen in Stereovision Systemen. Der zentrale Teil des Algorithmus ist die Entwicklung eines dualen Skalierungsfaktors, der es erlaubt, die perspektivischen Verzeichnungen sowohl direkt geometrisch als auch im Frequenzraum der Stereo-Bilder zu bewerten. Darüber hinaus trägt der Algorithmus zur eindeutigen Phasenbestimmung beim so genannten Phase-Wraparound bei und verbessert damit auch die Anwendung anderer auf der Phasenauswertung-basierender Methoden.

Der Algorithmus nutzt weiterhin die Amplituden- und Phaseninformation zur Schätzung von Tiefeninformation aus zweidimensionalen stereoskopischen Bildpaaren. Er berücksichtigt sowohl die Instabilitäten der Phase als auch die perspektivischen Verzeichnungen der Oberfläche (die Verkürzung in einer Ansicht). Diese Eigenschaften sind wichtig bei der Verwendung der Phaseninformation, um fehlerhafte Disparitätsschätzungen zu vermeiden. Im Weiteren wird anstelle des direkten Matchings der Intensitäten eine Gabor-Skalenraum-Erweiterung angewendet. Die Amplitudeninformation wird zum Entdecken "schwacher Punkte" im Frequenzraum verwendet, und nur verlässliche Phasenwerte bleiben für eine robuste Schätzung der Disparität übrig. Ein Vorteil dieses Algorithmus besteht darin, dass Disparitätswertschätzungen mit Subpixelgenauigkeit erhalten werden, ohne eine explizite Subpixelsignalrekonstruktion zu benötigen. Experimentelle Ergebnisse zeigen, dass der Verkürzungsfaktor seinen deutlichen Einfluss hat, wenn Objekte stark geneigt sind und sich in der Nähe der Kameras befinden.

Die Effizienz und Leistungsfähigkeit der vorgeschlagenen Algorithmen wurde auf der Grundlage der Analyse rektifizierter realer Bilder bestätigt. Die experimentellen Ergebnisse zeigen, dass der vorgeschlagene Algorithmus deutliche Vorteile hinsichtlich der Genauigkeit und Dichte der Disparitätsschätzungen im Vergleich zu alternativen Methoden aufweist.

Acknowledgement

All gratitude is due to ALLAH who guides me to bring forth to light this thesis.

I would like to extend my thanks and gratitude to Egypt for their financial support during my research without that, I was not able to work and search here.

I'm greatly indebted to my supervisor, Prof. Dr.-Ing. habil. Bernd Michaelis for being a consistent source of support and encouragement. His guidance and help have made my Ph.D. program a smooth and enjoyable one.

I am also grateful to Prof. Dr. rer. nat. Peter Hauptman, and Prof. Dr.-Ing. Abbas Omar, University of Magdeburg, for kindly agreeing to be referees for this thesis in spite of their hectic schedules.

Also I would like to thank Dr. Ayoub Al-Hamadi for his time in the many discussions that have returned to much interest

I always feel lucky to be with so many excellent researchers in Michaelis's group "AGMI". Thanks are due to all colleagues of my institute, who were always quite helpful during my stay.

I would like to thank my family; brothers and sisters provided much need support. I am deeply indebted to my father and my mother.

My most grateful thanks and loves to my wife (*Saheir*) and my kids (*Rawan, Rana and Ahmad*) who share all my live and who provided understanding and loving support during the difficult task of creating this dissertation. Again many thanks to my wife for her patience with me.

Sherif El-Etriby

Magdeburg, Germany

7.January.2008

Table of Contents

Dedications	i
Abstract	ii
Zusammenfassung	iv
Acknowledgement	vi
Table of Contents	vii
List of Figures	x
Notation	xvii
1 Introduction	1
1.1 Experimental Set-up and Calibration	2
1.1.1 Cameras	3
1.1.2 Light Projector	3
1.1.3 Calibration	3
1.2 Image Rectification and Depth Recovery	4
1.3 Stereo Vision Algorithms	4
1.3.1 Stereo Camera Geometry: Stereo Matching	4
1.3.2 Area-based Algorithm: Spatial Correlation	8
1.3.3 Feature-based	10
1.3.4 Phase-based	11
1.3.5 Problems In Stereovision: Ambiguity, Occlusion and Foreshortening	11
1.4 Related Work	14
2 Structured Lighting for Disparity Accuracy	16
2.1 Structured Light: Pattern Analysis	16
2.1.1 Generation of Structured Pattern:	19
2.2 Structured Light Effects	20
2.3 Reference Data: True Disparity	26

3	Linear Spatial Filters for Disparity Estimation	30
3.1	Local Analysis of Image Patches	31
3.1.1	Basis Filter Design: Gaussian Derivatives	31
3.1.2	Filter Optimization	33
3.1.3	Steerable Filters	34
3.1.4	Singular Value Decomposition (SVD)	37
3.1.5	Gaussian Filter Bank	38
3.2	Matching Filter Responses	44
3.3	Depth Discontinuities and Occluded Region	46
3.3.1	Adaptive Scale Selection	51
3.4	Experimental Results	53
4	Local Spatial Frequency Representation	59
4.1	Local Spatial Frequency	59
4.1.1	Spectrogram	61
4.1.2	Wigner Distribution	62
4.2	Gabor Filter	63
4.3	Gabor Scale-space Expansion: Scalogram	66
4.3.1	Phase-frequency Measurement	69
5	Phase-based Disparity Estimation	74
5.1	Phase-difference Based Method	74
5.1.1	Phase-difference as Disparity	74
5.2	Suggested Phase-difference Based Algorithm	78
5.2.1	Image Scalogram	79
5.2.2	Recognizing Singularity Neighborhood	79
5.2.3	Combined Algorithm for Estimating Disparity	82
5.3	Evaluation Methodology	87
5.4	Experimental results	92
5.4.1	Tuning Parameters	92
5.4.2	Ground-truth Evaluation	98
5.4.3	Comparison With Other Algorithms	98
6	Correspondence with Slanted Surface: Foreshortening Effects	104
6.1	Relating Disparity to Surface Angle	106
6.2	Foreshortening-factor using Image Parameters	108
6.3	Extending Phase-based Algorithm	110
6.4	Results	110
7	Summary and Conclusions	114
	Appendices:	119
A	Structured Light Pattern	120

Bibliography	128
Curriculum Vitae	135
Related Publications	137

List of Figures

1.1	Experimental setup for the 3-D measurement.	2
1.2	Experimental set-up devices.	3
1.3	A simple planar calibration target in different orientations. The calibration target have the size 40×40 cm with known geometrical dimensions of 28 marker points.	5
1.4	Visualization of finding the marker points geometrical determined on the calibration target.	5
1.5	Image rectification. Retinal planer are coplanar and parallel to the baseline.	6
1.6	Image before and after rectification. (a), (b) Images of 3-D scene parallel to the baseline obtained by a stereo camera system with converging optical axes. (c), (d) Rectification of the images so that corresponding epipolar lines fall on the same scanline in the images.	6
1.7	Normal case of stereo camera geometry with parallel optical axes. . .	7
1.8	Principle of spatial correlation windows: The patch window on the right is shifted pixel by pixel across a larger search window and the maximum cross-correlation to the reference on the left is found. . . .	10
1.9	Occlusion region. Stereo geometry with parallel axes for depth discontinuities.	12
1.10	Ambiguous match. Ambiguous matching occurs when a stereo method is unable to determine a unique correspondent for a particular pixel. .	13
1.11	Overhead view of the foreshortening model. The perspective distortion due to a slanted surface.	13
1.12	Unequal sampling in the stereo image pair due to the slanted surface.	14
2.1	The principle of structured lighting and stereovision. The system contain a projector (light source) and two cameras, a light pattern projected into the scene. The origin of the stereo camera coordinates system are $X_l Y_l Z_l$ and $X_r Y_r Z_r$. The considered illuminated object point $P(X, Y, Z)$ is formatted to a point $P_l(x_l, y_l)$ in the left image plane as well as point $P_r(x_r, y_r)$ in the right image plane.	17

2.2	Random-dot pattern 600×600 pixels. Gaussian white noise with mean 0.3 and variance 0.3.	21
2.3	Stripes binary pattern 600×600 , stripe either bright or dark has 8 pixels width.	22
2.4	Stripe pattern of much higher spatial frequency and corresponding image view.	24
2.5	Stripe pattern effect of a "Sinusoid" object. Three structured light of stripe pattern with varying period length ($0.8Cm$, $0.4Cm$, $0.2Cm$) projected onto the scene in (a1, a2, a3), a 3-D point cloud of the computed disparity map in (b1, b2, b3) and the 3-D surface reconstruction in (c1, c2, c3).	25
2.6	Data estimation based on a sequence of Gray coded patterns. (b)Example of the gray code bit planes, (c) greyscale images (bottom).	26
2.7	Search for the best correlation between the grey-scale values by shifting the pixel of right camera along the epipolar line.	28
2.8	Reference data from active stereo vision.	29
2.9	Examples of ground truth data used in the experimental results. (a) left image from the stereo pair, (b) the reference disparity map, (c) the 3-D points cloud and (d) the 3-D reconstructed surface.	29
3.1	Concept of the linear spatial filter (LSF) approach.	31
3.2	Left and right rectified stereo pair images "Doll images" to verify the steerable of Gaussian derivatives filters.	36
3.3	Set of steerable filters. The 2^{nd} derivative of Gaussian filters in different orientations and their responses images.	36
3.4	Concept of mathematical definition of SVD.	37
3.5	(Top) Orthonormal basis for the filters set shown in Fig. 3.3 as the output of SVD. (Middle) Plots of row-vector containing the variance of each column in the filter where x-axis is the filter length and the y-axis is the variance values. (Bottom) The filter response images of the corresponds filters.	39
3.6	Convolve a one column vector spatial filter, F_i , with a one column image patch, I	40
3.7	A linear transformation maps image patches I (where I is image patch $n \times 1$ column vector from the input image "Doll Image") to a vector of filter responses, $v = F^T I$	40
3.8	List of suggested linear spatial filters according to the filter set design as shown in Tab. 3.1, successive derivatives and rotations of Gaussian curves.	41

3.9	Set of filters. Four filters of the 1 st Gaussian derivative at four different orientations 0°, 30°, 60° and 90°, $\sigma = 4$ and filter size 31×31 pixels.	42
3.10	The adaptive filter set for the chosen filter set that are shown in Fig. 3.9.	42
3.11	Filter response images. (Top) is the filter response images obtained from convolving the adaptive set filters that are shown in Fig. 3.10 with the "Doll image" that is shown in Fig. 3.2. (Bottom) is a row vector of variances the filter response images.	43
3.12	The variance of each response image that reflects the amounts of information in each response image.	44
3.13	Matching the filter outputs by shifting the pixel of responses left image along the epipolar line.	46
3.14	Surface reconstruction for a flat surface. (a) Rectified stereo pair images with stripe pattern projector, (b) left computed disparity maps, (c) right computed disparity map, (d) the ground truth disparity (Reference disparity), (e) the 3-D point clouds, (f) the 3-D surface reconstruction.	47
3.15	(a) The left and right images of stereo pair "Hand", (b) the reference disparity map, (c) the computed disparity map, (d) the 3-D point cloud, and (e) the 3-D surface reconstruction.	48
3.16	Stereo geometry and visible projection regions. (Left) Stereo geometry with parallel axes for depth discontinuities view. (Right) Visibility map. The zeros mark the regions detected to be visible only from one of the two views.	49
3.17	Occlusion interpretation. The disparity map of the left image (left column), detecting occlusion map of the right image (right column).	50
3.18	Visibility map. The white area in the left and right visible maps mark the regions that contain no correct disparity value.	52
3.19	Scale selection in the depth discontinuities scene.	53
3.20	Convolving a left and right images with a set of linear spatial filters. (a) The set of nine Gaussian filters, at filter size 31×31 . (b) Stereo pair images of "Cylinder" object. (c) The response images of the left image.	55
3.21	Initial estimated disparity map for a stereo pair image of a "Cylinder" object computed from the suggested linear spatial filter (LSF) approach.	55
3.22	Progress the iteration of our algorithm to follow the results develop. The scan line from the computed disparity shown in the blue line and the reference scan line shown the red line in each iteration.	56

3.23	Comparison of the correspondence lines profiles from the computed disparity (CompDisp), the disparity from the Area-based method (Area-Corr) and the reference disparity (RefDisp).	57
3.24	Two Examples of the final computed results of "Doll object" in left and "Cylinder object" in right. The stereo pairs (top), the reference disparity map and the refined computed disparity map (middle). The color error map and the 3-D point cloud for surface reconstruction (bottom).	58
3.25	Refined disparity estimate to the same stereo pair shown in Fig. 3.18. Left and right computed disparity and its ground truth disparity map.	58
4.1	Local Spatial Frequency representation. (a) The one dimensional synthetic signal, a low frequency sin wave embedded in a high frequency wave. (b) The Local Spatial Frequency plot (Scalogram phase) associates each sample point with its proper frequency. (c) The Discrete Fourier Transform, which has two peaks, one for each frequency. (d) The Spectrogram representation, using a fixed kernel window without explicit knowledge of the original input signal's analytic form.	61
4.2	Gabor filter composition with tuning frequencies $\omega = 0.05$; (a) sinusoidal wave of particular frequency, (b) a Gaussian kernel, (c) the corresponding Gabor filter in a space domain, (d) Gabor filter in a frequency domain.	63
4.3	Gabor filter examples with constant values ($m = 4$ and $\sigma_f = 1/6$) but different values of the tuning frequencies (from top to bottom the ω equals 0.04, 0.0333, 0.0286 and 0.025). First column shows the Gabor filter in the space representation, and the second column shows their plots of frequency response.	65
4.4	Gaussian distribution with mean μ , and three standard deviations of the mean (three on each side).	66
4.5	Gabor filter example with the tuning frequencies $\omega = 1/10$, number of wavelength $m=4$ and $\sigma_f = 1/6$. The Real-Imaginary pair, and the Magnitude-Phase pair are presented from left to right.	67
4.6	Magnitude and Phase Scalogram of a one dimensional sinusoidal input signal. The horizontal axis of the Scalogram correspond directly with the signal's horizontal axis. The vertical axes of the Scalogram correspond to a frequency scale with low frequency at the bottom.	68
4.7	Discontinues view and its scalogram behavior: (a) Signal to be analyzed, (b) the magnitude scalogram, and (c) is its phase scalogram. In the phase plot the arrows mark vertical lines of constant phase that occur at step transitions in the signal.	70

4.8	Instantaneous frequency analysis for two signals. The signal with fixed frequency (left) and signal with varying frequency (right).	72
5.1	Estimating disparity between two images from phase difference and instantaneous frequency.	76
5.2	Disparity as a function of Phase difference. The disparity is a horizontal separation between the two images.	76
5.3	Concept of phase difference-base suggested algorithm. Besides on the phase and magnitude information of the rectified input stereo images using the measured-phase difference $\Delta_{\phi_{meas}}$ and the ideal-phase difference $\Delta_{\phi_{ideal}}$	78
5.4	Shape of the signal from a complete scan-line and the scalogram representation of this signal (both magnitude and phase).	80
5.5	Detection of singularity neighborhoods. The scalogram response (magnitude and phase components) of a line-profile from figure 5.4, with $1 \leq \lambda \leq 20$. The two plots in left column are the magnitude $\rho(x, \mu)$ and phase $\phi(x, \lambda)$. In the middle column, the level contours of the magnitude on top and phase in bottom. Right column is the local frequency and the phases as a function of spatial position at $\lambda = 7$	82
5.6	Characteristic behavior of level phase and magnitude contours in singularity neighborhoods. A) the small ellipsoidal contour marks location at which $\frac{\partial \phi(x, \lambda)}{\partial x} = 0$, the other ellipsoids contours illustrate the regions detected by the constraints. B) the level magnitude contours near singularity that appear like a parabola.	83
5.7	Application of stability constraint. Detection of singularity neighborhoods by the combine constrain 5.7 for $\tau = 0.06$ and $4 \leq \lambda \leq 20$. A) level phase contours and the region which detected shown in the labeled regions. B) The correspondence Magnitude contours that achieve the constraint.	83
5.8	Sample usable magnitude that correspond to the stable phase values at pixel number (75,150) from the pair images. only the values inside the dotted rectangles are used in the computation.	85
5.9	Ideal phase difference as a function of candidate disparity illustrate the phase wraparound problem. Each graph has 30 sample points spaced linearly a long the x-axis.	86
5.10	Experimental results of the Doll image pair with disparity range (0..15). The accuracy and the smoothness of the computed result are observed from the comparison, that is because the proposed method is optimally captures both local orientation and frequency information from the input image.	88

5.11	Principle of the quality measurement for disparity map without ground truth.	91
5.12	The choice of frequency. The computed disparity at $\omega = 1/10, 1/30, 1/50$, while two other parameters (m, σ_f) are kept fixed at $m = 4, \sigma_f = 1/6$, disparity range $[0...15]$. the stereo images composed of $N=256$ points (as pike, like one dimensional stereo signal). The figures are translated on the Y axis for visualization.	94
5.13	Effect of σ_f . The transition steepness is marked by the arrows.	95
5.14	Effect of a varying number of wavelengths m on the disparity estimates. All of the filters in the figure have constant frequency $\omega = 1/10$ and $\sigma_f = 1/6$	96
5.15	Line profile of computed disparity from our adaptive filter compared with the ground truth in different values of σ_f	97
5.16	Resulting disparity maps from the Middlebury stereo data set. (top row) Left image; (second row) Ground truth [1]; (third row) Disparity maps generated from our method; (forth row) Error map contain bad pixels (in black), correct pixels (in white) and excluded occlusions (in gray);(bottom row) is the occluded region assigned in the white areas.	99
5.17	Percentage of bad pixels in $B_{\tilde{O}}, B_{ho}$ and B_{no} of various algorithms.	101
5.18	3-D surface reconstruction for cylinder stereo pair images.	102
5.19	Multi object scene. (a) left image from the stereo pair, (b) reference disparity map, (c) computed disparity, (d) is the visualization of Error map where the correct pixels shown in gray-level from 32 to 64, the invalid pixels in blue and the bad matching pixels in magenta, (e) shows visualization of 3-D surface reconstruction, (f) lines profile comparison from the computed (CompDisp) and the reference disparity (TrueDisp).	103
6.1	Foreshortening model. X_S is the distance from the original point, O_S , which exactly in front of the left camera to the point S on the surface. X_L and X_R are the left and right segment indices of the image for surface distance X_S ; the surface is viewed at a sharp angle θ	105
6.2	Unequal projection lengths for the object surface acquired in a slanted view.	105
6.3	Stereo pair images show more compressed in one image due to perspective foreshortening. The projected segment X_L is shorter than X_R	106
6.4	Foreshortening effects in the scalogram representation.	109
6.5	Results for slanted object. (a) are Flat surface rotated by angles $0^\circ, 20^\circ$ and 45° . (b) are the ground truth disparity maps (reference disparity). (c) are the 3-D surface reconstruction rotated for a good visualization.	112

6.6	Ground truth compare against the computed disparity maps, for the center scanline of the flat surface at the various rotation angle; 0° , 20° and 45° and disparity range (1...6) pixels.	113
A.1	Random-dot pattern 600×600 , the brightness value in the range 0...255.	121
A.2	A combined Gaussian white noise in default parameters; zero mean noise and 0.01 variance with a stripe line of width 8 pixels.	122
A.3	Checker-board pattern, with a basic structure alternating bright and dark squares.	123
A.4	Sinusoidal pattern with a period length 80 pixels.	124
A.5	Sawtooth pattern with a period 80 pixels	125
A.6	Sinusoidal-Sawtooth pattern with a period 80 pixels	126
A.7	Color stripes pattern generated as a sequence of light stripes with brighter colors, $C_1, B, C_2, B, \dots, B, C_N$, a black stripe B is inserted between each C_i and C_{i+1} to increase the intensity variation between adjacent light stripes.	127

Notation

Symbol	Description
$3 - D$	Three dimensional.
<i>ABW</i>	Automatisierung Bildverarbeitung (ABW LCD-640); Light stripe projector with a maximum resolution of 640×640 lines.
<i>B</i>	Base line. The distance between the centers of projection.
<i>f</i>	Focal length. Intrinsic camera parameter associated with the pinhole camera model.
<i>d</i>	Disparity range: The range of candidate disparities from d_{min} to d_{max} which the actual disparity value was located within the crossed disparity range.
<i>disp</i>	Disparity candidate.
ω	Frequency, cycles per unit distance.
λ	Wavelengths in units of pixels.
<i>LSF</i>	Linear Spatial Filter.
<i>FRV</i>	Filter response vector.
<i>SVD</i>	Singular value decomposition.
<i>FT</i>	Fourier Transform.
<i>STFT</i>	Short-Time Fourier Transform.
<i>NCC</i>	Normalize cross correlation.
<i>SSD</i>	Sum of squared difference
<i>CCD</i>	Charge-coupled device (CCD), Camera is a sensor for recording images.
<i>BitDepth</i>	Integer indicating the number of bits per pixel.
ϕ	Phase. ϕ_l and ϕ_r denoted as the phases of the left and right images.
ρ	Magnitude <i>or</i> Amplitude.

Symbol	Description
m	Wavelengths per window: The Gaussian envelope of a Gabor filter will be truncated outside this many wavelengths. This determines the window size.
ρ_f	Sigma fraction: Given a fixed window size (m), the σ parameter of Gaussian will be this fraction of the window.
*	Convolution operation.
$SP(x, \omega)$	The spectrogram representation of the power of frequency ω at the point x .
$WD(x, \omega)$	Wigner distribution.
$G(x, \omega)$	Gabor filter centered at the origin for tuning frequency ω and point x .
S_R	Scalogram representation.
$\Delta_{\phi_{ideal}}$	Ideal-phase difference.
$\Delta_{\phi_{meas}}$	measured-phase difference.
EF	Evaluation function that computes a quantitative agreement between the two sets of phase differences ($\Delta_{\phi_{ideal}}$ and $\Delta_{\phi_{meas}}$).
ADM	Absolute difference module. The smallest difference between phases.
$BadM$	Percentage of bad matching pixels between the computed disparity and the ground truth disparity.
$RMSE$	Root mean squared error.
$PSNR$	Peak signal to noise ratio.
$B_{\tilde{0}}$	Non occluded region.
B_{ho}	Half occluded region.
B_{no}	Near discontinuities.
ADE	Absolute disparity error.
MRE	Mean relative error.
PRE	Percentage relative error.
$DevAE$	Deviation of the average error.
Ψ	Foreshortening factor.

Chapter 1

Introduction

Stereo matching is a technique used to extract depth from a pair of images that was taken from a slightly different view points. In stereovision, the most challenging problem for disparity estimate and surface reconstruction is to find corresponding image points. This problem is called the *corresponding problem* in stereo matching algorithm. Given a pair of corresponding pixels, the distance between the two cameras (baseline) and their orientation, it is easy to apply triangulation to find the distance to the point in world coordinates represented by those pixels (i.e. reconstruct the 3-D information). In order to do that, our task is to focus on finding the vector offset between corresponding points, which is called the *Disparity*.

Common computation algorithms have been proposed for disparity measurements. These algorithms differ from one another in matching primitives, the density of the results, the accuracy of the estimates and the underlying computation time. In general, stereo algorithms can be classified in three classes: area-based, feature-based and frequency-based stereos. Section 1.3 will explain these algorithms in more details.

The frequency-based, well known as phase-difference-based technique, has become a widespread method for depth and optical flow estimation, because of its superior performance and better theoretical grounding.

In frequency-based technique, the original signal is transformed to Fourier space and some parts of the transformed signal (Magnitude and Phase component) are used to estimate the disparity [2], [3], [4]. In our thesis we attempt to overcome some of the long-standing problems in stereovision by using two algorithms [5], [6], [7], [8]. The first algorithm [5], [6] is based on the outputs of linear spatial filters tuned to a range of orientations and scales. Convolving the left and right images with a bank of linear filters tuned to a number of different orientations and scales (2-D oriented derivative of Gaussian filters). The responses of these filters at a given point constitute a vector that characterizes the local structure of the image patch. This algorithm allow us to address scale space selection and half occlusion problems, but their solution require an initial disparity map that is computed without the benefit of models for these problems.

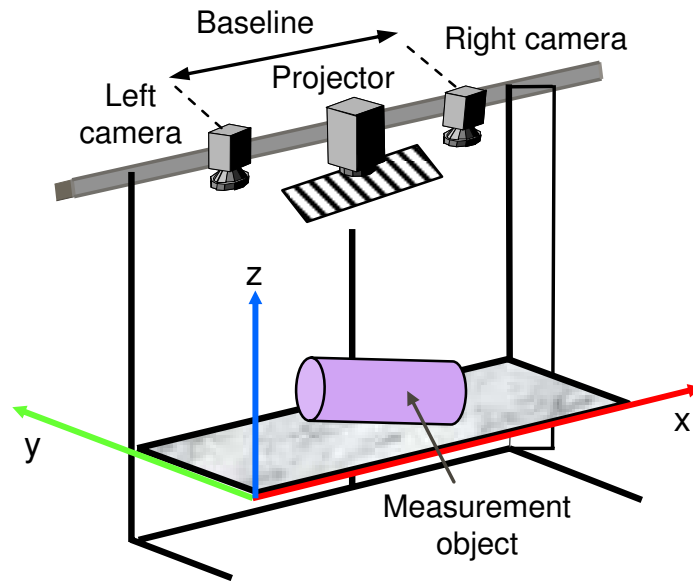


Figure 1.1: Experimental setup for the 3-D measurement.

The second algorithm [7], [8] (Phase-difference based algorithm) based on the convolution of the stereo image pairs with Gabor-scale space expansion and caching the outputs in a local spatial frequency representation (scalogram). Gabor filter has two parameters; the width and the tuning frequency. In order to optimize the performance, these parameters have been chosen in accordance to the characteristics of the visual signal. In our thesis, we propose an automatic technique to locally adapt the filter parameters to the input signal. In the first, we analyze the performance of the phase-difference-based technique for disparity estimation with respect to the choice of the Gabor filter parameters. Afterwards, a novel technique is introduced that reduces phase nonlinearity by means of an adaptive mechanism for the tuning frequency. The performance improvement that is produced by the adaptive filter is demonstrated using different types of images.

1.1 Experimental Set-up and Calibration

The experimental set-up for 3-D analysis is shown in Fig.1.1. Image sequences or stereo pair images are acquired by two or more cameras with parallel optical axes. In our method, the surface coordinates of the measurement object are determined from the images recorded by two cameras and a projector device situated between the cameras. The projector emits a structured lighting patterns into the scene where the projected light pattern illuminates the scene and the object. The reconstruction of the 3-D surfaces in my research is based on the usage of a one shot taken from the

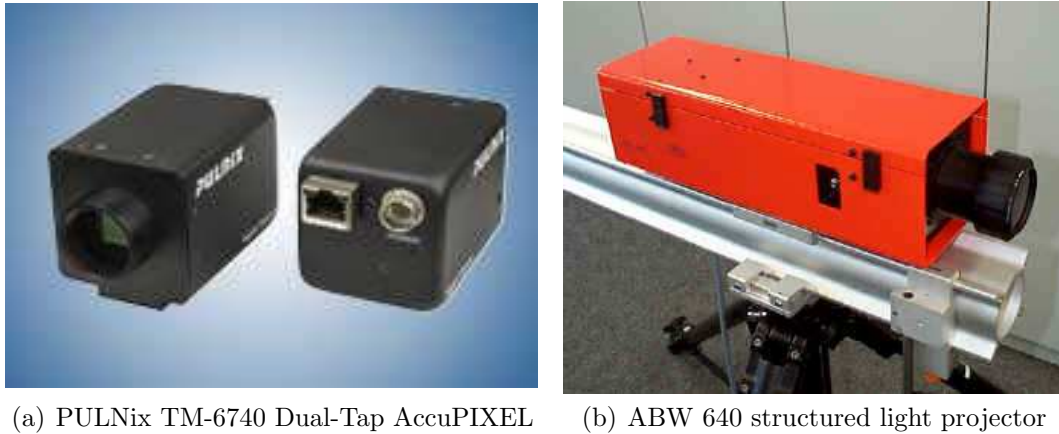


Figure 1.2: Experimental set-up devices.

stereo camera of an object illuminated with a stripe pattern.

1.1.1 Cameras

A pair of PULNix Dual-Tap AccuPiXEL cameras were used for high resolution acquisition. The cameras have about 50 *cm* measured baseline separation between the two cameras in our experimental results.

The PULNix TM-6740 Gigabit Ethernet camera is a compact, VGA format monochrome progressive scan CCD camera in the AccuPiXEL Dual-Tap family, utilizing the GigE interface. Based on a high-quality Kodak KAI-0340D 1/3" image sensor, this camera's most outstanding feature is in its ability to capture 200 frames per second at full resolution and up to 3205 fps in partial scan and binning modes. The cameras 640(*H*) × 480(*V*) resolution, as shown in Fig. 1.2(a) .

1.1.2 Light Projector

Our system consists of also a ABW LCD-640 (Automatisierung BildverArbeitung) stripe projector with a maximum resolution of 640 × 640 lines that can be used to generate arbitrary line patterns to project the Structured Light pattern into the scene, as shown in Fig. 1.2(b).

1.1.3 Calibration

The calibration of the system by the standard method (e.g. bundle adjustment) [9] is a basic requirement before acquiring the data images. Calibration is implemented by using a calibration-target with marker points of known geometrical dimensions, as depicted in Fig. 1.3. The calibration process can be summarized in briefly as follows: First, we find all ellipses centers pints for all the marker points in the target figure. These centers are numerics as we can see the image in Fig.1.4. Second, calculate the

approximate values of the camera parameters. Finally, the approximate values are improved by bundle block compensation [9]. As a result, the extrinsic parameters; position of the projection center, the angle with respect to the object coordinate system as well as the intrinsic parameters; effective focal length, scale factor, principle points and distortion parameters of the cameras are available [9] [10] [11] .

1.2 Image Rectification and Depth Recovery

A common problem in all stereovision tasks is the correspondence problem. To simplify the search for the corresponds points in the stereo images, or to simplify the search in image structures representing the same world structure, images are usually rectified. The result are a pair of images where corresponding points lie on the same horizontal line. This way limiting the search region [12]. Image rectification is a transformation process used to project multiple images onto a common image surface. It is used to correct a distorted image into a standard coordinate system.

If the two camera axes are not parallel, their associated epipolar lines are not parallel to the scan lines [13], [9]. The process of rectification for a pair of images transforms the original pair of image planes to another pair in a way that the resulting epipolar lines are parallel and equal along the new scan lines. Rectification is depicted in Fig.1.5. Here C_L and C_R are the camera optical centers, U_1 and U_2 the original image planes and V_1 and V_2 refer to the rectified image planes. The condition of parallel and equal epipolar lines requires planes V_1 and V_2 to lie on the same plane. A point P is projected to image points O_l and O_r on the same scan line in the rectified planes.

Rectifying the images is one of the important processes in the correspondence analysis. It reduces the problem of the canonical case when the corresponding mutually visible points are in the same row. An example of a rectified wide baseline stereo pair with sample epipolar lines is shown in Fig. 1.6. In order to do that, the rectification program reads two camera parameters files which are the output of camera calibration and the number of unrectified pair images. Caching the outputs of the rectified images and the rectifying camera parameters i.e. in the normal case (parallel optical axes).

1.3 Stereo Vision Algorithms

1.3.1 Stereo Camera Geometry: Stereo Matching

The normal case stereo camera geometry has optical axes that are parallel and are normal to the baseline which lead to the notation of disparity which is often used in stereo literatures [1] [14]. This geometry is depicted in Fig.1.7, the image of an object point $P(X, Y, Z)$ is formatted at (x_l, y) in the left image and at (x_r, y) in the right

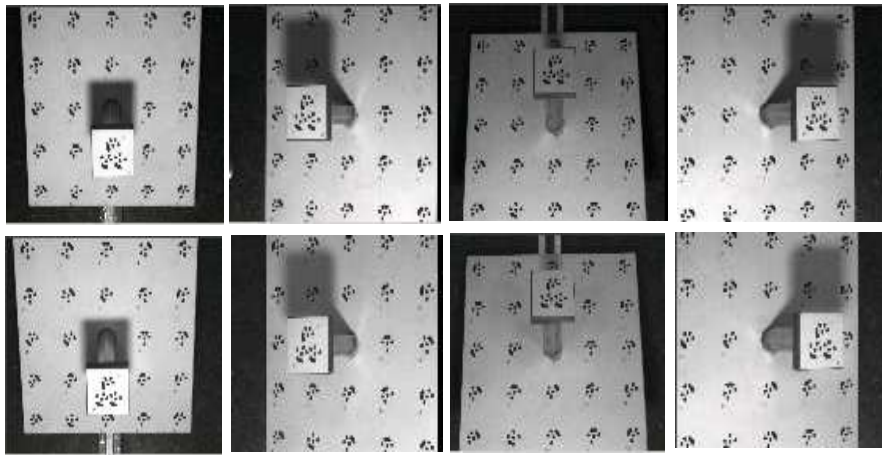


Figure 1.3: A simple planar calibration target in different orientations. The calibration target has the size 40×40 cm with known geometrical dimensions of 28 marker points.

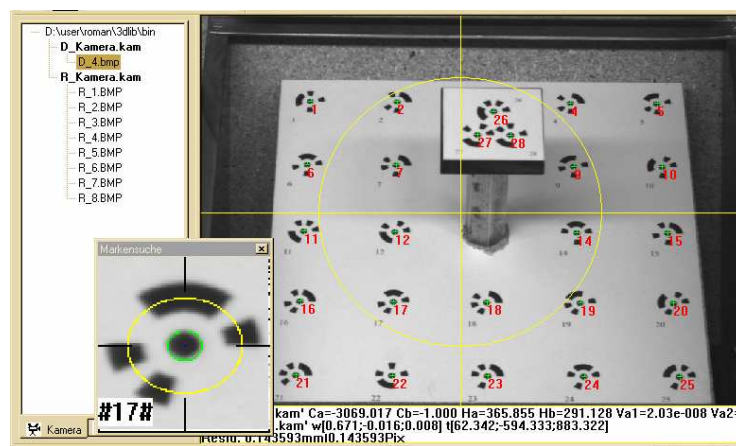


Figure 1.4: Visualization of finding the marker points geometrical determined on the calibration target.

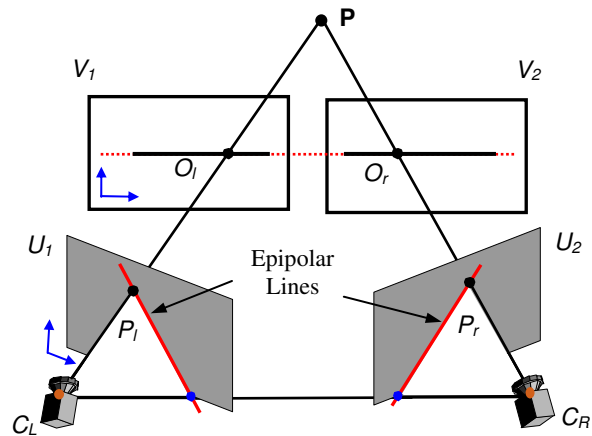


Figure 1.5: Image rectification. Retinal planes are coplanar and parallel to the baseline.

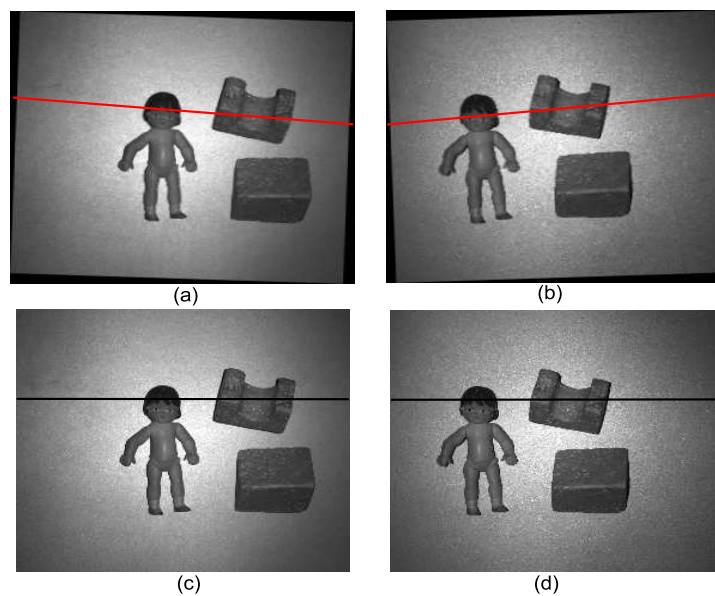


Figure 1.6: Image before and after rectification. (a), (b) Images of 3-D scene parallel to the baseline obtained by a stereo camera system with converging optical axes. (c), (d) Rectification of the images so that corresponding epipolar lines fall on the same scanline in the images.

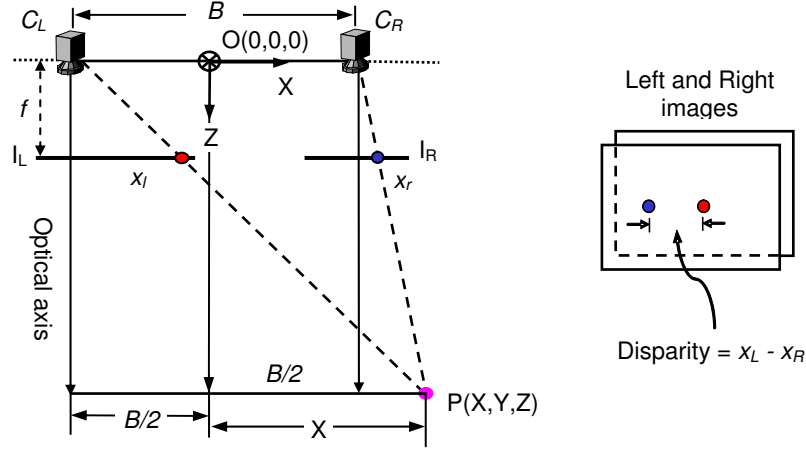


Figure 1.7: Normal case of stereo camera geometry with parallel optical axes.

image. The stereo-camera coordinate system is considered to be midway between the left and right camera coordinate system.

Note that, the baseline B and the focal length f are positive numbers, the parameter Z (the distance between the baseline and the object point) is a positive coordinate and x_l, x_r are coordinates that may be positive or negative. The relation between the 3-D world point when measured with respect to the left camera coordinate system and the stereo-camera coordinate system can be written by the following equation.

From similar triangles in Fig. 1.7, for the left camera, we have

$$\frac{x_l}{f} = \frac{X + B/2}{Z} \quad (1.1)$$

and similarly for the right camera, we have

$$\frac{x_r}{f} = \frac{X - B/2}{Z} \quad (1.2)$$

this can be written as

$$X + B/2 = Z \frac{x_l}{f} \quad (1.3)$$

and similarly for the right camera, we have

$$X - B/2 = Z \frac{x_r}{f} \quad (1.4)$$

removing X from Eq.1.3 and Eq.1.4, we can obtain

$$Z = \frac{B \cdot f}{(x_l - x_r)} \quad (1.5)$$

Then, we obtain the canonical expression relating horizontal disparity ($x_l - x_r$) to depth Z (distance between the base line and the object point)

$$(x_l - x_r) = \frac{B \cdot f}{Z} \quad (1.6)$$

the disparity is the entity that has be measured for each pair of corresponding points to infer 3-D coordinates or surface point position of visible scene points from a given stereo image pair. Eq. 1.6 which is presented also in [14], [15] gives pointwise disparity only; we will show how to extend this description to surfaces at arbitrary angles.

The disparity can be estimated using various methods which can be organized into the three categories:

- Area-based: Correlation-based, [16], [17], [18], [19], [20], [21]
- Feature-based, [22], [23], [24]
- Phase-based, [3], [4], [25], [2], [8], [26]

These methods differ from one another in the matching primitives, the density of the results, the accuracy of the estimation and the underlying computation time. The most reported signal step methods in the literature are: feature-based, correlation-based and phase-based methods.

In the following subsections, we will describe the advantages and the disadvantages for each method and the common problems in stereovision.

1.3.2 Area-based Algorithm: Spatial Correlation

As we mentioned above in section 1.3.1, stereo matching is performed with two images which satisfy an epipolar line constraint. With the assumption of pin-hole camera, the relation equation is derived as Eq. 1.5. Common computational approaches include feature-based, area-based and phase-based methods. All these methods have their intrinsic problems as they are caused by the assumptions inherent in these algorithms.

The area-based stereo algorithm has widely used for practical stereo vision systems. The matching process is applied to the intensity profiles of the two images. They have a very good matching accuracy in surface areas with continuously smooth disparities. So, the appropriate size selection of a rectangular window is of importance. Generally, if the size of matching areas is determined as large as possible, enough intensity variations are included and then matching accuracy is improved in continuously smooth areas. However, large deformation of matching areas results from projective distortion and smoothing occurs in the vicinity of depth discontinuity. Therefore, existing area-based stereo algorithms tend to generate bad match results in depth discontinuity.

An accurate disparity estimation is indispensable to automated reconstruction because 3-D measurement depends on the disparity. Difficult problems in stereo matching include matching ambiguities. In addition, intensity differences of corresponding points become larger due to the projective distortion in occlusion boundary. The intensity value of stereo images is defined as I_l and I_r respectively. The relations of intensity value can be expressed by Eq. 1.7;

$$I_l(x, y) = I_r(x + d, y) \quad (1.7)$$

where $d(x, y)$ is disparity function at a pixel position.

In area-based methods, normalized cross correlation (NCC) and sum of squared difference (SSD) are mainly used as matching criteria [27]. Correlation is widely used as an effective similarity measure in matching tasks. However, traditional correlation based matching methods are limited to the short baseline case.

The matching can be used efficient phase correlation [28] or simply the mean normalized Cross correlation. The algorithm compute similarity error for every pixel in the image by taking a fixed window in the left image and shifting it along the epipolar line in the right image, see Fig. 1.8. The correlation coefficient between the reference window and the search window is calculated as follows:

$$NCC(x, y) = \frac{\sum_{j=1}^N \sum_{i=1}^M [l(i, j) - \bar{l}] \cdot [r(x + i, y + j) - \overline{r(x, y)}]}{\sqrt{\sum_{j=1}^N \sum_{i=1}^M [l(i, j) - \bar{l}]^2} \cdot \sqrt{\sum_{j=1}^N \sum_{i=1}^M [r(x + i, y + j) - \overline{r(x, y)}]^2}} \quad (1.8)$$

In Eq. 1.8 [29] x and y are coordinates of the correlation window in the left image. The summation are performed over all pixels in the correlation window. l and r are pixels from the left and right correlation window respectively, \bar{l} and $\overline{r(x, y)}$ are their mean values over the correlation window. NCC is preferable since it is invariant to linear brightness and contrast variations between the perfect matching windows. The value of Eq. 1.8) is between -1 and +1 and a larger value indicates more similarity between windows.

In Table 1.1 the summarized pseudo-code for a simple stereo matching algorithm. A central problem in area-based method is to find the optimal size of the support region. If the region is too small, a wrong match might be found due to ambiguities and noise. If the region is too big, it can no longer be matched as a whole due to the foreshortening and occlusion with the result of lost detail and blurring of object boundaries.

The main problem in such correlation based stereo is that the results at depth discontinuities and occlusion boundaries are generally quite unreliable, Fig. 1.9 shows the occlusion region in the discontinuities view. Also, make use of the same sized

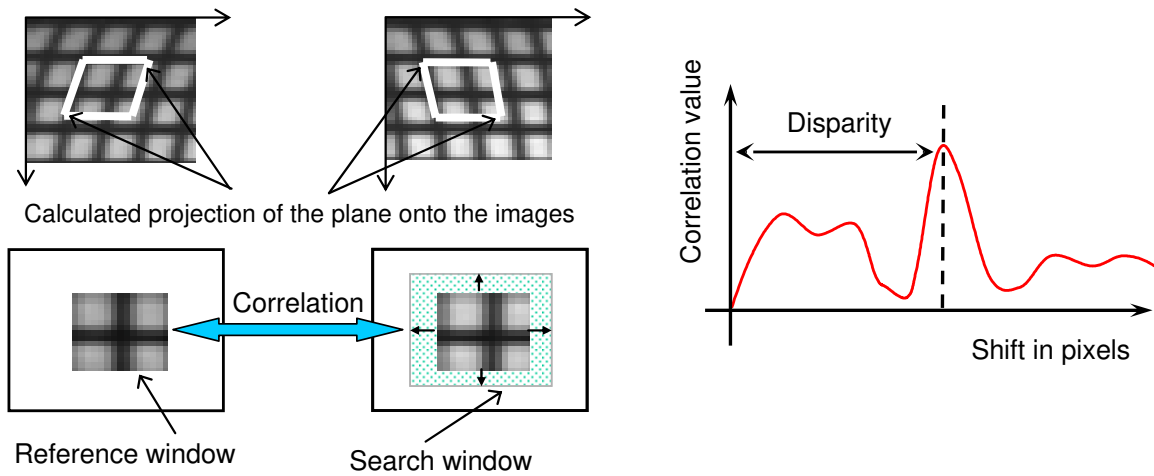


Figure 1.8: Principle of spatial correlation windows: The patch window on the right is shifted pixel by pixel across a larger search window and the maximum cross-correlation to the reference on the left is found.

Table 1.1: **Pseudo code for Stereo Matching Algorithm**

Given: A pair of grayscale images, disparity range,

- 1: Let $I_l(x_l, y_l)$ and $I_r(x_r, y_r)$ are the left and right images
- 2: Let $EF(x_l, y_l, x_r, c_r)$ be the similarity function
- 3: **for** each pixel (x_l, y_l) in L_l
- 4: **for** each possible match (x_r, y_r) in L_r
- 5: Compute $EF(x_l, y_l, x_r, y_r)$
- 6: **end for**
- 7: Let $\Delta(x_l, y_l)$ be the choice value which yields the min/max value of EF
- 8: **end for**

matching windows in both images which mean that the objects are assumed to be basically planar and are viewed from the front that will be demonstrated and tried to over come these outliers in our method.

1.3.3 Feature-based

The feature detect methods that are based on finding distinct features in images such as lines, angles, corners ... etc. A list of features is stored for each image and then the lists are compared to determine correspondence.

The matching stage operates only on these extracted image features. This method uses less calculations than area-based, but does not have unique correspondences between left and right images due to the projection onto a subspace. Moreover, since correspondence is only establishing at a small number of pixels, the resulting

depth map will be very sparse. Thus while feature-based matching may be useful in applications where the features are known to be visible and only sparse disparities are required, it is not so useful for extracting dense depth maps.

1.3.4 Phase-based

In the phase-based methods, the local phase frequency of the horizontal 1-D signals are extracted. In our method, an adaptive Gabor filters convolved with the scanline form the entire images. Phase and magnitude can be calculated from these filter responses and then the disparity can be computed.

The local phase has a number of interesting in-variance and equi-variance properties; local phase estimation are invariant to signal energy [30], local phase estimation and spatial position are equi-variant and the spatial derivative of local phase estimates is equi-variant with the spatial frequency, (the phase derivative is called local or instantaneous frequency). More details can be found in chapters four and five.

1.3.5 Problems In Stereovision: Ambiguity, Occlusion and Foreshortening

The input to any binocular stereo system is a pair of images. The task is to match primitives of the two images, thereby solving the correspondence problem. The depth of objects in the scene can then be determined. There are several choice of matching primitives and the performance of disparity estimation is affected by different sources of problems [31]. We distinguish between;

1. Error sources that are due to the basic assumptions on which the computational procedure of stereovision is found.
2. Sources that are due to the approximations introduced in the mathematical procedure.

Regarding the first issue, note that stereovision is based on the assumption that a pair of stereo images are locally related by a one dimensional shift. Since the stereo images are two dimensional scenes taken from slightly different views, we expect; presence of occlusion, difference in scale and also the probability of the miss matching.

In spite of the strong physical constraints available to stereo methods, many problems remain in their implementation. While the mapping from disparity to depth is well understood, the automatic extraction of disparity is still subject to error. A complete review of most stereo matching problems is beyond the scope of this work.

- **Occlusion Region:** In the scene that contains more than one level surface, this situation is depicted in Fig. 1.9 region C located in the near surface and regions A, B, D and E located in the far surface. There are often regions visible

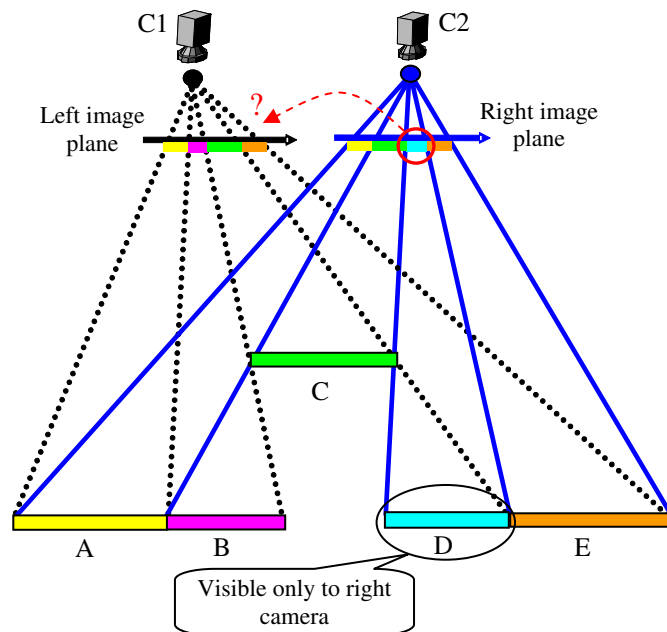


Figure 1.9: Occlusion region. Stereo geometry with parallel axes for depth discontinuities.

to one camera but not the other. These regions called occluded regions. In this region there is no corresponding point in the other view and recovered disparity estimates appear as error.

- Ambiguous Matching:** As we can observe from the image in Fig. 1.10, The grid pattern in the background is completely repetitive. So, the matching for a point in that region is difficult. The SSD (sum of squared differences) over a small window is one of the simplest and most effective measures of image matching. For a particular point in the base image, a small image window is cropped around it and as it is slid along the epipolar line of other images, the SSD values are computed for each disparity value. Such SSD values with respect to disparity for a single stereo image pair is shown as the right plot of Fig. 1.10. As expected, it has multiple minimums and matching is ambiguous. Because the camera separation and object-to-camera distance are restricted, the object's perimeter will not differ significantly in the two images. The object boundary descriptions for each image can be compared prior to the matching procedure to ensure minimal global correspondence. This check may reveal ambiguous matching, i.e. situations for which a local point or area in one image may match equally well with a number of points or areas in the other image [32], [33]. If the perimeters differ by more than 20 percent, then the object is considered "failed to match".

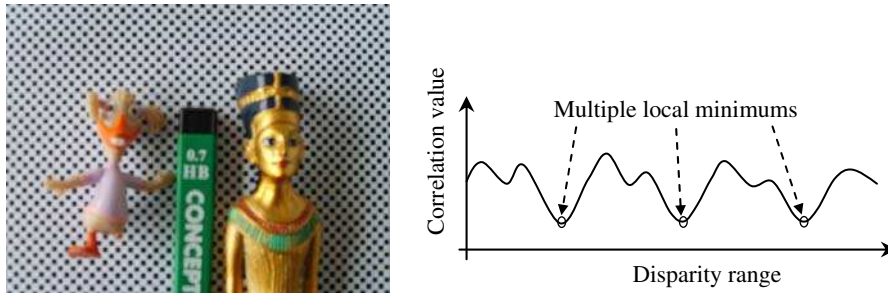


Figure 1.10: Ambiguous match. Ambiguous matching occurs when a stereo method is unable to determine a unique correspondent for a particular pixel.

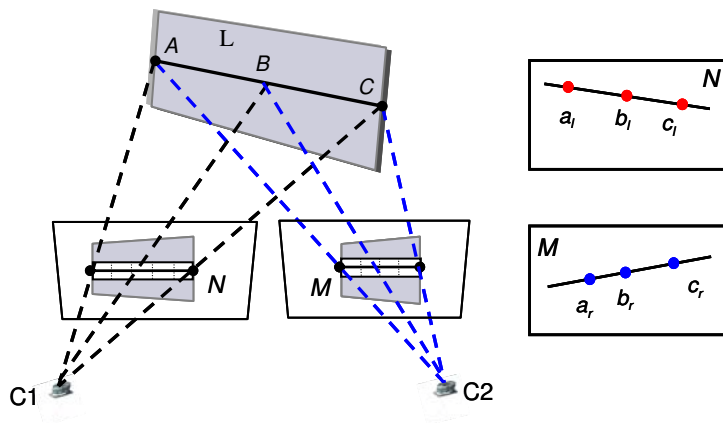


Figure 1.11: Overhead view of the foreshortening model. The perspective distortion due to a slanted surface.

- Unequal Projection Lengths (*Foreshortening*):** The object surface is viewed head-on in both images of a stereo pair. In other situation the object surfaces are acquired in a slanted view, therefore, they appear more compressed in one image, due to perspective foreshortening. Fig. 1.11 illustrate this problem where a projection line segment L_1 from left camera and a projection line segment L_2 from right camera have a different number of pixels. When a surface has a textured appearance, this effect makes matching its two images very difficult, since its appearance differs so much between the two images.

In Chapter 6, we develop a model of perspective foreshortening and provide a foreshortening factor that allow us to reason foreshortening effects in 3-D world coordinates and 2-D image coordinates.

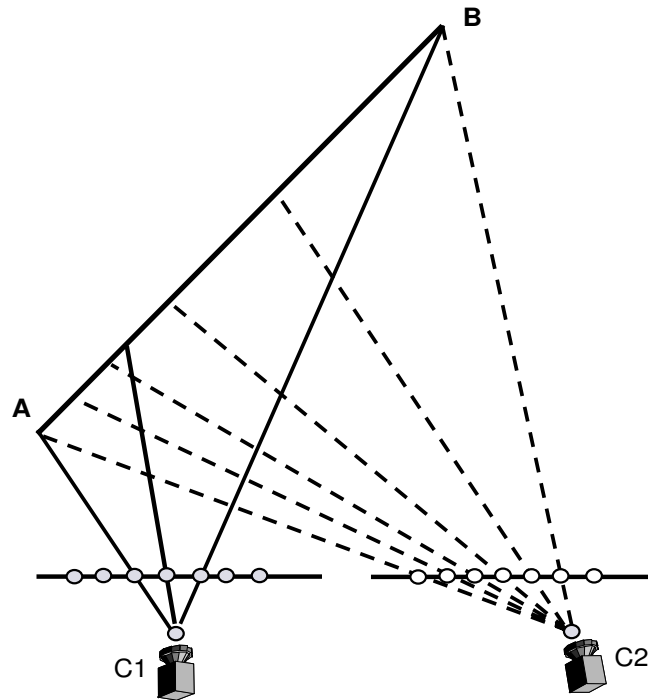


Figure 1.12: Unequal sampling in the stereo image pair due to the slanted surface.

- **Unequal Sampling (Unequal intensity):** Since a line segment has different projection lengths in the two images, therefore its intensity functions also sample different, as we can see in Fig. 1.12.

These topics will be addressed in the forthcoming chapters.

1.4 Related Work

Over the last few years, a large number of techniques have been developed [25], [2], [34], [7], [35], [36], [26], [37] etc. Scharstein and Szeliski [35] have provided an exhaustive review and comparison of stereo correspondence algorithms. Recently, several works proposed frequency-based (phase-based) techniques to disparity estimation. The simplest method is to minimize the matching error within rectangular windows of fixed size, Kanade T., [16] presented an adaptive window method to reduce the effect of projective distortion. His method employs a statistical model of the disparity distribution within a window, by evaluating the local variation of the intensity and disparity. This method can select an appropriate windows size and estimate disparity with the lowest uncertainty for each pixel of an image. Better approaches utilize multiple windows [17], [18], shiftable windows [19], [20], or predicted windows [21], all of which give performance improvement at discontinuities. Kim [38] presented a

new stereo matching algorithm based on window warping technique in hierarchical matching process to balance the perspective distortions. Falkenhagen [39] presented a hierarchical block-based approach by considering neighborhood constraints and by estimating hierarchically on a multi-resolution image pyramid of a stereoscopic image pair.

Y. Ohta [22] used derivatives to find edges on each scanline, and then link edges between different scanlines to provide global constraints.

In contrast to correlation-based or feature-based approaches which suffer from high computational load, the approach of the frequency domain is characterized by the original scan line in the 2-D image which is transformed to a frequency domain. Afterwards, disparity can be calculated directly from local phase differences. Furthermore, phase-based algorithms provide inherent subpixel accuracy, as they yield information directly from the phase difference. When estimating disparity in the frequency-based algorithm, not only phase but also magnitude information is available to have an accurate disparity estimation. A combined approach, which takes phase and magnitude into account, should be promising for fast and robust disparity estimation. The quality of reconstruction is directly dependent on the quality of the disparity map. One of the first examples of the phase-based algorithms is the Kuglin-Hines method [40], which utilizes the phase shift theorem of the Fourier transform. The method proposes a model based on the inverse Fourier transform of the phase difference between two images. Singer [34] was one of the first researchers who proposed the use of the phase difference between two local filter responses in order to compute the disparities of the different object in the two stereo images. Jones and Malik [41] applied local spatial frequency and use the magnitude of the filter responses at each pixel as matching features, but used an affine transformation matrix in the spatial domain without providing a description of its effect in frequency domain.

Fleet et al. [3] have shown that the results can be improved by dividing the local spatial frequency of the response instead of the tuning filter frequency. The same authors [30] also provide additional phase measurements constraints to address the phase wraparound problem. As a matter of fact; surface normal is often very tiled with respect to the optical axes of camera, the projected stereo images has a projected distortion which appear more compressed in one image. This effect makes matching of its two images very difficult, since its appearance differs between the two images. This leads to confusing results from area-based approaches, because the visible areas vary between the two images. Weng [4] used the windowed Fourier phase, which is the multiple-window using the Kuglin-Hines algorithm. The window size is allowed to vary so that the measurements can be more localized. Ahlvers [25], [2] presented a new approach for combining magnitude and phase information in Fast Fourier Transform (FFT)-based algorithms.

Chapter 2

Structured Lighting for Disparity Accuracy

2.1 Structured Light: Pattern Analysis

Stereovision is an attractive and widely used method, but, it is rather limited to reconstruct the 3-D surface map, due to the correspondence problem. The correspondence problem can be reduced using a method based on the structured light concept.

Among all the 3-D reconstruction methods, active stereovision systems are becoming increasingly important. An active stereovision system is composed of one light source and one or more camera. The light source can be a spatial light modulator or a video projector. Each camera is fixed in front of the object at different positions.

The projection of light patterns into a scene is called "Structured lighting". The light patterns are projected onto the objects which lie in the field of view of the camera. The simplest and best recognizable light patterns are light spots and strips. The distance of an object to the camera can be determined through analyzing the observed light patterns in the images. The active manipulation of the scene by using light patterns simplifies the 3-D reconstruction task [15].

To calculate the 3-D coordinates of object surface points, the projector projects a certain light pattern onto the measured object, and each camera acquires an image of the highlighted object. The 2-D coordinates of a surface element in the images are extracted respectively. Fig.2.1, shows the principle of structured lighting and stereovision system.

The most popular approach in structured lighting is to use a single stripe light. The advantage of this approach is that it greatly simplifies the matching problem. But this approach has the drawback that only one single line of 3-D data points can be obtained with each image shot at one time. The most widely used method of this approach proposed by [42], he use only one single camera together with a calibrated light source, which projects a known pattern of light on the measuring scene. It is required to find the correspondence between the projected pattern and the imaged

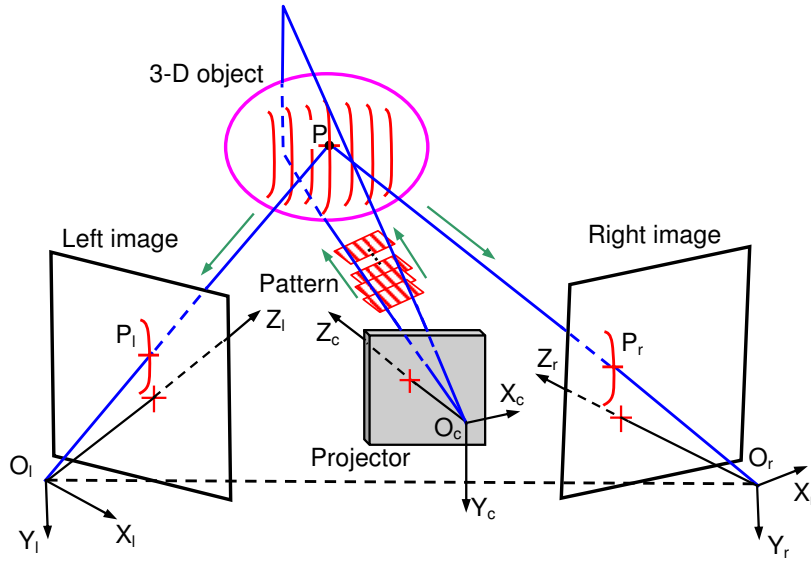


Figure 2.1: The principle of structured lighting and stereovision. The system contains a projector (light source) and two cameras, a light pattern projected into the scene. The origin of the stereo camera coordinate systems are $X_l Y_l Z_l$ and $X_r Y_r Z_r$. The considered illuminated object point $P(X, Y, Z)$ is formatted to a point $P_l(x_l, y_l)$ in the left image plane as well as point $P_r(x_r, y_r)$ in the right image plane.

one. It is also needed to calibrate both the camera and the light source.

In order to speed up the acquisition of 3-D range data, we adopt to use a multiple-stripe light pattern projected onto the measurement object. Two cameras are placed at different view points to capture the left and right images (stereo pair images). Thus increasing the local discriminability of each pixel and facilitating matching process. Once the correspondence problem is accurately solved, the 3-D range data can be computed by using triangulation. Therefore, by using more than one camera, we can replace the more difficult problem of lighting-to-image correspondence by an easier problem of image-to-image stereo correspondence. Furthermore, we do not need to calibrate the position and orientation for each of the projected light in 3-D space, i.e. cameras are the only devices to be calibrated.

As it is mentioned, most of the proposed structured light techniques are based on the projection of regular patterns on the measuring scene. If a single light dot or a slit line is projected on the scene, then there is no correspondence problem to be solved, but whole the scene has to be scanned to obtain the 3-D map. Shirai et al., [43] proposed a slit line projection to recognize curvilinear object. Agin [44] generalized this idea to recognize curvilinear objects. In 1986 Yamamoto [45] proposed a half plane illumination system instead of a slit line. Recently, there are also some authors, Sato et al. [46] as well as Kemmotsu [47], who use a method based on the

projection of two or three slit lines with different orientation and position in the 3-D coordinates system. Asada [48] proposed to use a pattern made by a set of vertical parallel and equidistant stripe lines. Wang [49] have extend Asada's idea with the sequential projection of two orthogonal strips patterns. Chen [50] have successfully developed and implement a new method for 3-D range data acquisition by combining color structured lighting and stereovision. Pattern of color stripes is projected onto the objects while each camera acquires image. Edge segments are extracted from the acquired stereo images pair, and then used for finding the correct stereo correspondence. Kang [51] proposed a 4-camera system to recover dense stereo range data from each set of images, a sinusoidally varying pattern project onto the scene to enhance local intensity discriminability.

All these method allow us to obtain 3-D information from the geometric constraint propagation, especially from the epipolar constraint. Recently, structured light technique has increased important in most of stereo vision methods. This technique is based on a unique codification of each token of light projected on the scene. This technique is basically known as coded structured light. In general, by looking at the light projected, patterns can classified into:

- *Binary*: Any of the two coordinates (x_p, y_p) of the projected point can only have one of two possible values, which are coded with 0 and 1 respectively. This binary value normally represents opacity and transparency on the object.
- *Grey level*: Each pattern point can have an associated grey value, which represents the transparency (or opacity) level of the point against the projected light. Since the information is coded as a grey light level, normally two steps are necessary in order to find out 3-D information. First, we have to obtain an image of the scene illuminated with the same light intensity for each point (without coding). Second, we must obtain the reference light needed to cancel the surface reflection effect, which depends on the kind of surfaces where the light is reflected. This limitation means that the pattern has to be also classified as a static pattern.
- *Color*: Each pattern point has to be associated with a hue value. In order to use the color constancy property, the hue values used have to be quite different from each other. The main goal is to get an efficient and accurate segmentation. Since the system projects color on the scene, its use limited to a neutral color scene, as highly saturated color objects can produce losing of pattern regions in the segmentation step and posterior de-codification. Even pale colored objects may produce a colored frequency shifting as a result of the intrinsic color of the measuring objects. Obviously, the discrepancy suffered by the color captured by the camera with respect to the one projected by the pattern, rather complicates the segmentation step.

2.1.1 Generation of Structured Pattern:

Structured lighting simplifies the task by increasing the engineering prerequisites, moreover we can obtain rich and highly specific image features. Various shapes of light patterns exist, e.g. spot patterns, stripe patterns, checker-board, sinusoidal or color patterns. The position, orientation and shape of the light patterns can be changed or remain static during the image acquisition process.

Structured lighting is especially applied in those fields where automated three-dimensional measurement of an object has to be carried out with high precision, of course, this class of techniques is restricted to environments allowing the active projection and detection of light patterns. Even outdoor scenes are recoverable without introduction artificial illumination. For example, sunlight in conjunction with a thin pole produces a strip of shadow on the objects which yields equivalent information.

The motivation for using structured lighting is based on the expectation of the precise detection of the projected light patterns in the acquired images, the correspondence analysis is considerably simplified in both images. This is because patterns are projected onto the scenes to facilitate image point correspondence.

Random-dot Pattern: This pattern is generated as a Gaussian white noise with mean = 0.3 and variance = 0.3. The primary advantage of this pattern is that it is stochastically unique everywhere, and the corresponding pointed can be detected reliably [52]. In practice, we are interested to analyze the generated patterns, Fig. 2.2 (a) shows a square wave to illustrate the concept of spatial frequency resolution where spatial frequency refers to how rapidly the brightness signal is changing in space. Fig.2.2(b) shows the generated random-dot pattern 600×600 pixels. A central block 16×16 of the original pattern is used to test the correlation performance of the pattern by using the normalized cross correlation (NCC), recalling Eq. 1.8, where NCC is invariant to linear brightness and contrast variations between the perfect matching windows. The middle row in the same figure show the auto-correlation maps, the normalized cross correlation for the whole pattern is shown in graph (d), while the correlation of the template window is shown in graph (e) and the 3-D shaded surface plot of the correlation shown in (f), as a results the template was embedded at the center coordinate (300, 300) and the NCC found at the same point (300, 300). In other view, bottom row in Fig. 2.2 give us the behavior of the pattern in the frequency domain. A central line profile taken from the original pattern is represented in a discrete Fourier transform (bottom-left), as well as, the scalogram representation shown in bottom-right. More details on the scalogram representation will be explained in chapter four. Briefly, to generate the scalogram an adaptive Gabor filter $G(x, \omega)$ convolve with a one dimensional input row from the original pattern, caching the output in a two dimensional matrix scalogram, magnitude and phase, as shown in the bottom-middle and the bottom-right respectively. This representation is very useful for image matching especially with frequency-based algorithms. This is

because the scalogram-phase provides an actual detection to the localization of the varying frequency.

Binary Encoded Stripe Pattern: For the binary encoded light stripe projection, a set of light planes is projected onto the examined objects at the same time, these light patterns lead to a unique code for each plane, see Fig.2.3. Therefore, 2^n light planes are uniquely encodable by using n patterns, i.e. taking n images. Then a large number of images has to be generated for the object. The complete analysis of this pattern illustrated in Fig.2.3, and for an implementation example Fig.2.5 shows a gray code image sequence. The light planes were vertically projected onto the object.

Sawtooth Pattern: another type on the projected patterns is the sawtooth pattern with a period length bright illuminated on one side and half brightly on the other. Obviously, the maximum surface depth discontinuity to be measured is limited to the chosen period. In each period the illumination intensity is constant along the x-axis but increasing along the y-axis.

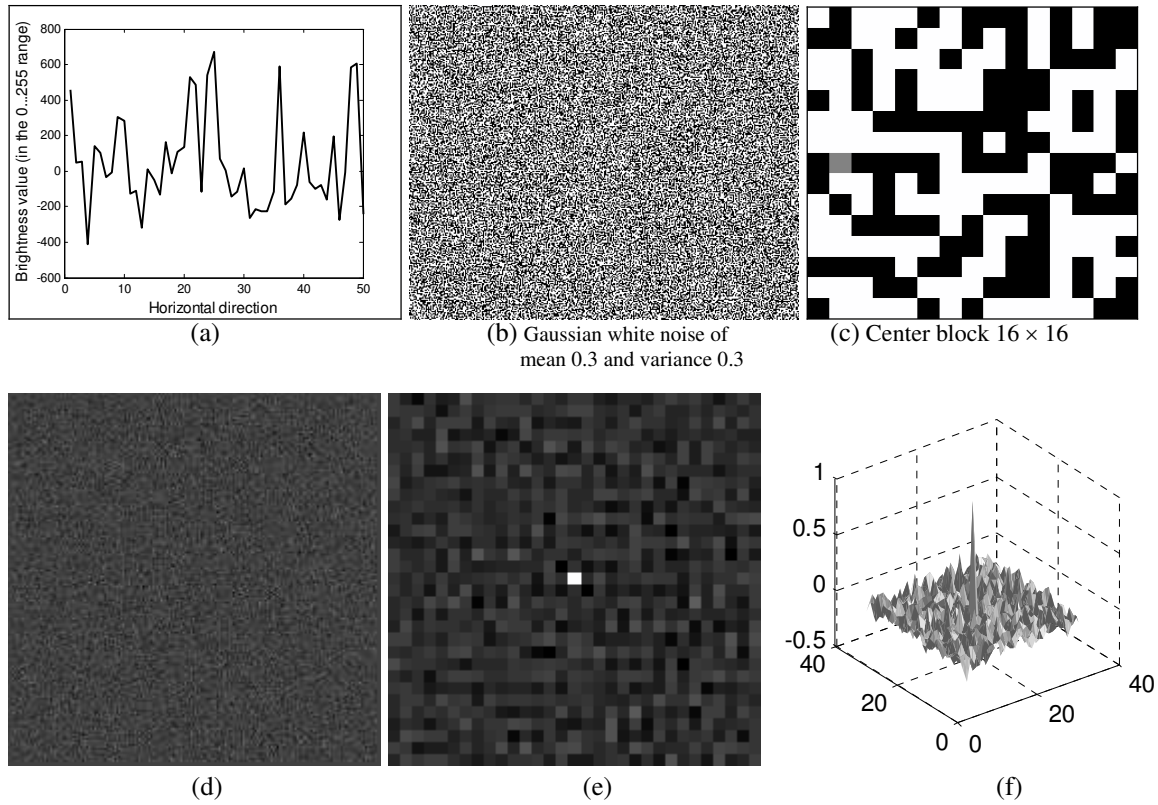
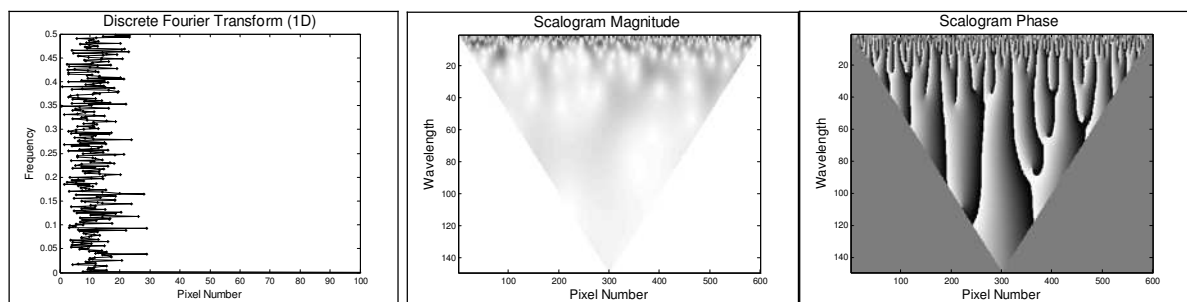
Color Stripes Pattern: For generating a color structured pattern, e.g. color stripes pattern, Chen [50] generate a sequence of light stripes with brighter colors, called C_1, C_2, \dots, C_N , in order to increase the intensity variation between adjacent light stripes, a black stripe B is inserted between each C_i and C_{i+1} , ($i = 1, 2, \dots, N - 1$). Hence the final generated pattern is $C_1, B, C_2, B, \dots, B, C_N$. Fig.A.7 in appendix A shows the color stripes pattern, wherever, it can be seen that the pattern has large variations between adjacent light stripes, and the correlation between any two consecutive color sequences is quite small.

Appendix A, "Structured Lighting", contains the complete figures for the patterns analyzed, Sawtooth pattern, Sinusoidal pattern, checker-board pattern, Color stripes pattern etc, Fig. A.1 - A.7. In each figure, graph (b) show the originally 600×600 structured lighting pattern, a central block 16×16 depicted in graph (c), a distribution map of auto-correlation values calculated by NCC function using a 16×16 pixel region in the center part of the pattern depicted in graph (d), (e) and (f). The bottom row in each figure represent the frequency representation of a central scanline from the origin pattern, the discrete Fourier Transform on the left graph, and the *scalogram* representation (its Magnitude and Phase) on the right graph.

2.2 Structured Light Effects

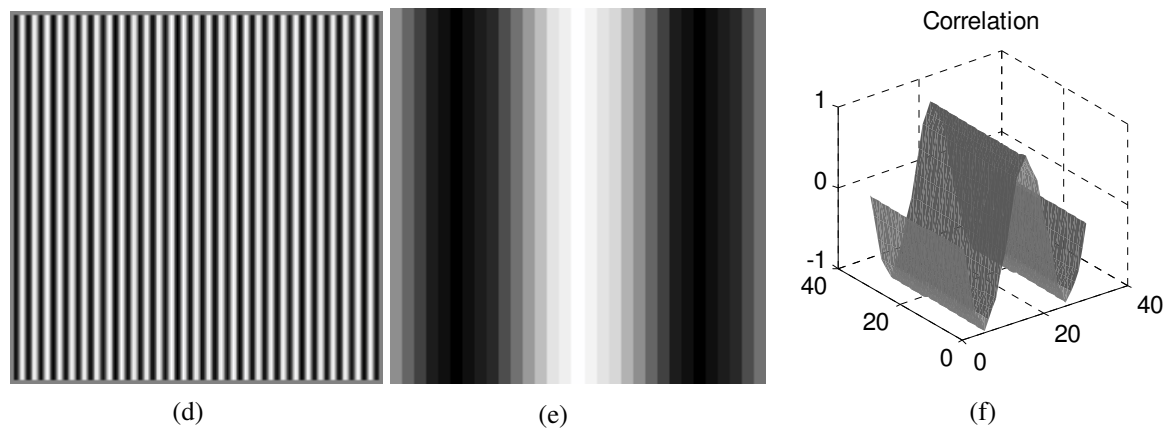
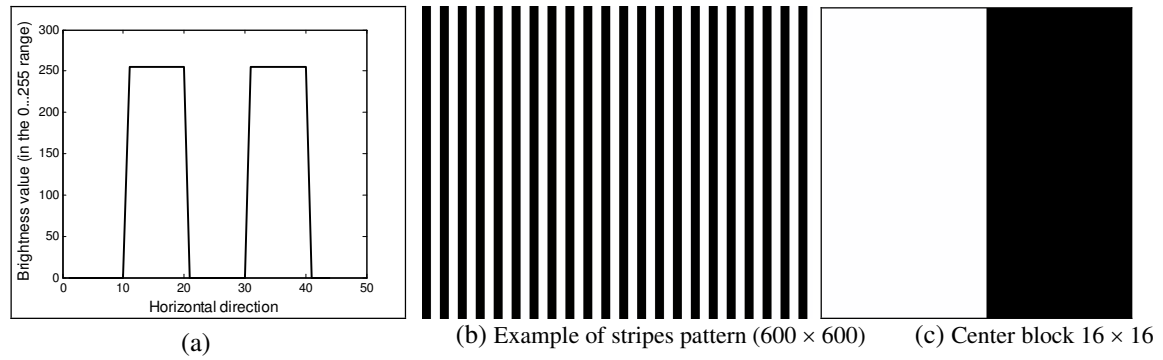
This section, deals with the effect of structured lighting patterns on the accuracy of disparities estimates, as well as, the 3-D reconstruction surface.

The location (x_i, y_i) of a pixel p_i in the image constrains the 3-D location of the corresponding object point $P_i(X_i, Y_i, Z_i)$ to a certain sub-space in the scene. Therefore, by using a disparity between each corresponding points and known camera geometry parameters, the 3-D position (X_i, Y_i, Z_i) is obtained.

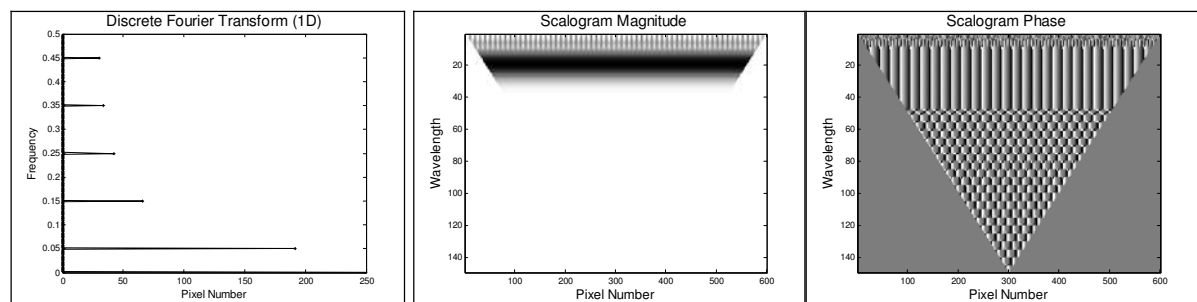
Auto-correlation map calculated by NCC for center block (16×16)

Spatial-frequency response for a one line profile from the pattern. The discrete Fourier transform and the scalogram (Magnitude and Phase) representation calculated as Eq. 3.5

Figure 2.2: Random-dot pattern 600×600 pixels. Gaussian white noise with mean 0.3 and variance 0.3.



Auto-correlation map calculated by NCC for center block (16 x 16)



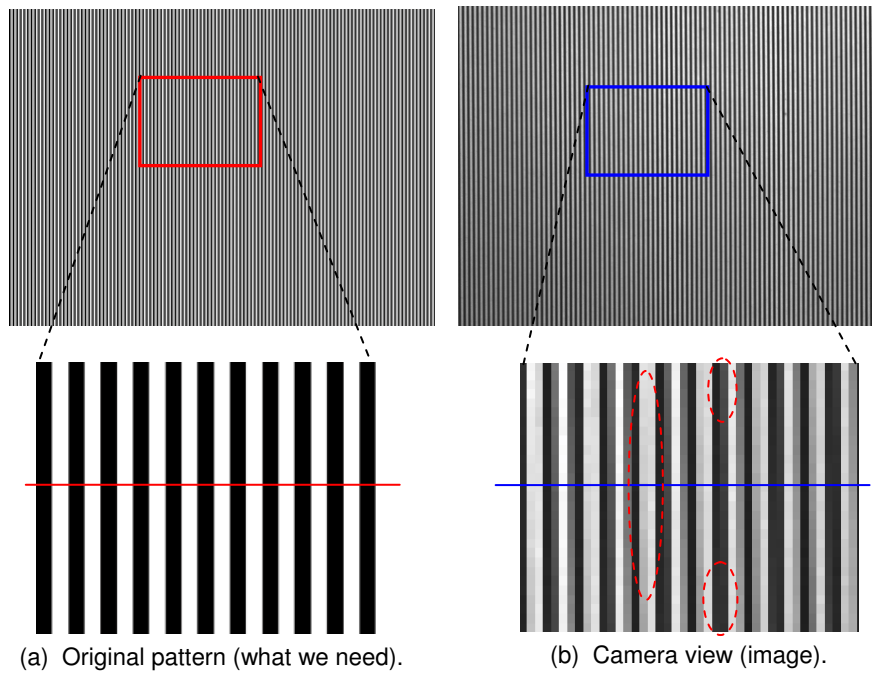
Spatial-frequency response for a one line profile from the pattern. The discrete Fourier transform and the scalogram (Magnitude and Phase) representation calculated as Eq. 3.5

Figure 2.3: Stripes binary pattern 600×600 , stripe either bright or dark has 8 pixels width.

Most stereo matching algorithms can not compute correct dense depth maps for homogenous image regions. This is due to the ambiguity of image values inside these regions. The ambiguity can be reduced by adding a synthetic texture to the scene. Of course, the accuracy is depending on the projected pattern and the stereo vision technique, when two or more parts of an image pair are similar in appearance. This can happen when a repetitive pattern like checker-board is present. A part of the pattern in one image might seem to match several parts in the other. It can not determine a unique correspondent for a particular pixel, as we can observe from the results of the auto-correlation map in Appendix A, Fig. A.3(e, f), where we see more than one peak for the particular pixel. Consider two images of a front planer checker-board where the checker-board occupies the entire fields of view; the alignment of the squares can not be determined from the images alone. Exactly, these type of images will cause problems for stereo matches and the effect can be seen in disparity estimates that are wildly inaccurate.

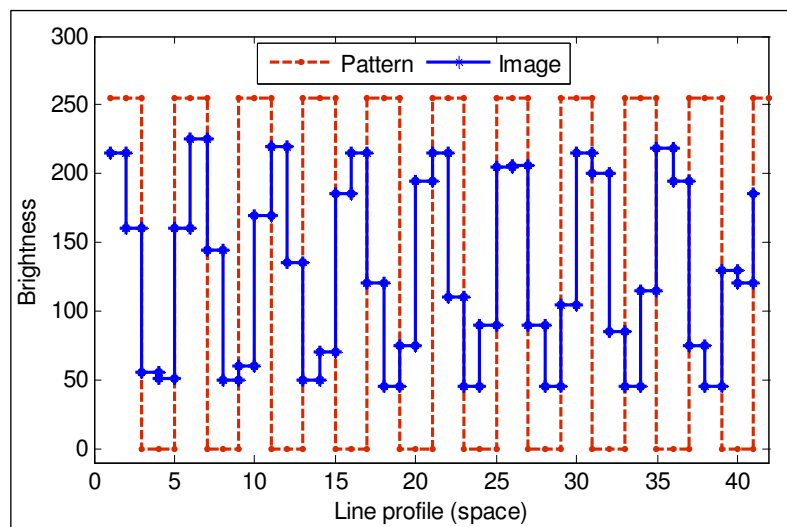
As it is well known, the use of structured lighting simplifies the difficult matching problem. But in general we can say that there are different advantages and drawbacks. For example, the use of stripe pattern introduces a "phase ambiguity" during stereo matching causing an inaccurate depth to be calculated. However, this can be overcome with the appropriate adjustment to the wavelength of each stripe in the sine grating. This gives each "stripe" a unique width that improves stereo matching accuracy, especially in regions of a low texture difference. Fig. 2.4 explores the relation between periodic spacing (spatial frequency) of the strips and the actual camera view (image).

In order to understand the concept of spatial frequency resolution, we need to define exactly what we mean by resolution. Resolution has to do with the ability to separate two adjacent strips, if we can see two adjacent pixels as being separate, then we can resolve the two. Generally, If strips in two cycles (a cycle is one complete change in the signal; in the stripe pattern we need at least two stripes one dark and one light for a cycle) appear as one or can not exist a light separate between them, then we can not resolve the two and this cause miss matching when solving the corresponding problem. That means if we increase the frequency, the stripes get closer and closer together, until they start blending together. Fig. 2.4(a) shows the higher frequencies of the original pattern. The camera view (image) of this pattern is shown in Fig. 2.4(b). However a rotated camera will catch a distorted view of the stripe pattern, with the amount of deviation or disparity, that is proportional to the object's height above the reference plane. We can observe in the marker areas that the strips are blended together. Fig. 2.4(c) illustrates the space diagram of two sample line profile from the original pattern and its corresponding image view. Since correct behavior is the brightness transformation from dark to light and from light to dark and so on, while in the image view the brightness does not have the symmetric behavior.



(a) Original pattern (what we need).

(b) Camera view (image).



(c) Space diagram for two lines profile from the original pattern (on dashed line) and camera image (on solid line).

Figure 2.4: Stripe pattern of much higher spatial frequency and corresponding image view.

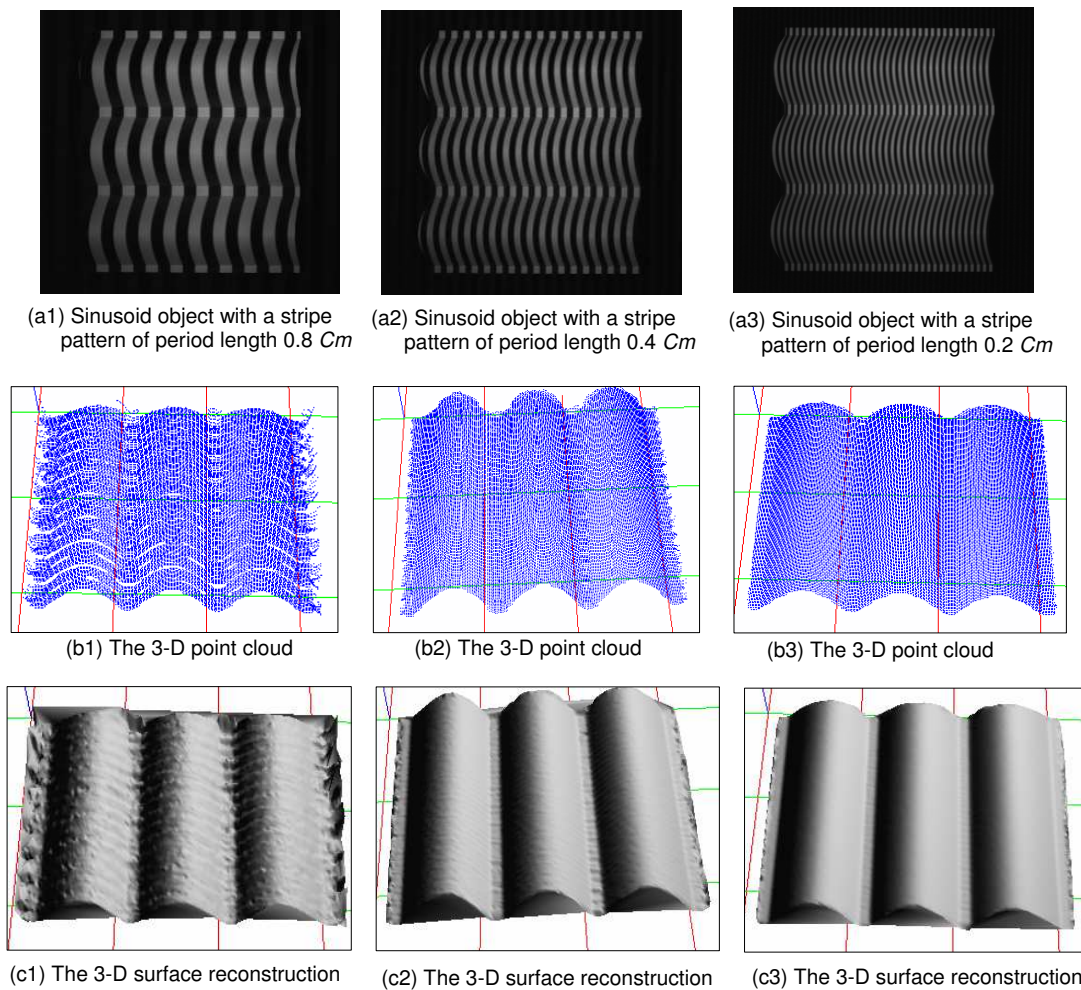


Figure 2.5: Stripe pattern effect of a "Sinusoid" object. Three structured light of stripe pattern with varying period length (0.8 Cm , 0.4 Cm , 0.2 Cm) projected onto the scene in (a1, a2, a3), a 3-D point cloud of the computed disparity map in (b1, b2, b3) and the 3-D surface reconstruction in (c1, c2, c3).

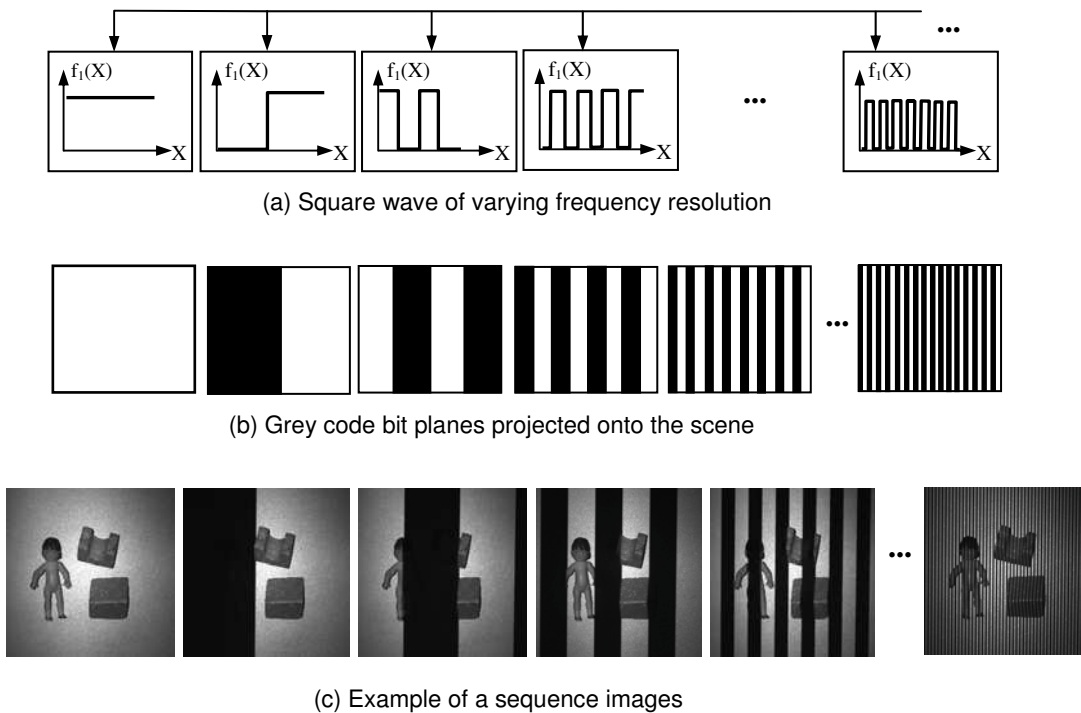


Figure 2.6: Data estimation based on a sequence of Gray coded patterns. (b) Example of the gray code bit planes, (c) graycode images (bottom).

The effect of varying the frequency (period length) of the stripe pattern is depicted in Fig.2.5. The figure shows a three left images of the stereo pares images taken from project the three different structured light patten with different period length (0.8 cm , 0.4 cm , and 0.2 cm) respectively in Fig. 2.5 (a). The 3-D point cloud of the computed disparity map in each case shown in Fig. 2.5 (b). Also the 3-D surface reconstruction shown Fig. 2.5 (c). It is clear to see the effect of varying the period length of the stripe pattern in the results. That is because in the low frequency stripe pattern (period length = 0.8 Cm) the given pixel in one view of the stereo images has multiple potential corresponds in the other view. Therefore the result has many of mismatching points (ambiguous problem).

2.3 Reference Data: True Disparity

In stereo vision algorithms, one of the interesting objectives is to quantitatively compare the performance of the existing stereo vision techniques. Unfortunately, the lack of benchmark datasets (Database images with their ground truth maps) makes it difficult. Consequently, we apply a fast active 3-D measurement of geometrical shapes by photogrammetry and structured lighting method that was presented in our

group [53], [54] to obtain a reference data set. Wherever, at our Lab in the university of Magdeburg, it is possible to make measurements based on photogrammetric method. The surface coordinates of the measurement object are determined from the images recorded by two cameras which are mounted at different locations. A projector situated between the cameras projects a texture onto the surface of the body to create sufficient grey-scale gradients. In this method [53], [54], instead of correlating area in two images, greyscale values of single pixel obtained from a number of different images taken in succession are correlated. Thus, it is possible to achieve higher spatial resolution. The technique uses a striped lighting method to identify projected light planes. The basic principle of the striped light method is the extraction of the height coordinates from the deformation of the projected lines caused by the measurement object.

The sequence of binary gray code images requires n patterns to distinguish among 2^n light planes. The light planes are projected onto the measurement object to create sufficient grey-scale gradients. (see Fig 2.6). Fig. 2.6 shows several consecutive patterns projected onto the measurement surface, the entirely illuminated scene is shown at the bottom, the images show the generated sequence from different projected light patterns shown in Fig. 2.6 (c).

Eleven images were used for codification. Nine of these are Grey code patterns which allow for distinguishing 512 light planes in an image. One fully illuminated image and one which is not are additionally taken. These are used to gather estimates of the on and off states for thresholding and subsequent binary grey code determination. Grey codes have the advantage that spatially adjacent profiles have codes which change by only one bit.

Measurement start with correlating only one pixel by using grey-scale value sequences of generating textures on the surface [54]. The principle of measurement consists in finding two pixels with approximately corresponding sequences of grey-scale values.

In addition, The algorithm for calculating a 3-D coordinate starts by selecting a pixel on left camera. The grey-scale values of this pixel from the n images produce a sequence of grey-scale values, see Fig. 2.7. The objective is to find the point with the corresponding grey-scale-value sequence on the other view. The desired corresponding point on right camera can only be located on the epipolar line. Therefore the necessary starting point for the procedure is also located on the epipolar line. A sequence of n grey-scale values from the n images is assigned to this point. For the search of the corresponding point, the correlation coefficient between the two grey-scale-value sequences of the two considered pixels is calculated (see Fig. 2.7). When the point is shifted along the epipolar line, the correlation coefficient changes. The gradient of this coefficient is determined in order to find the point which results in maximum correlation between the grey-scale-value sequences. This is the desired point, which corresponds with the point selected on left camera.

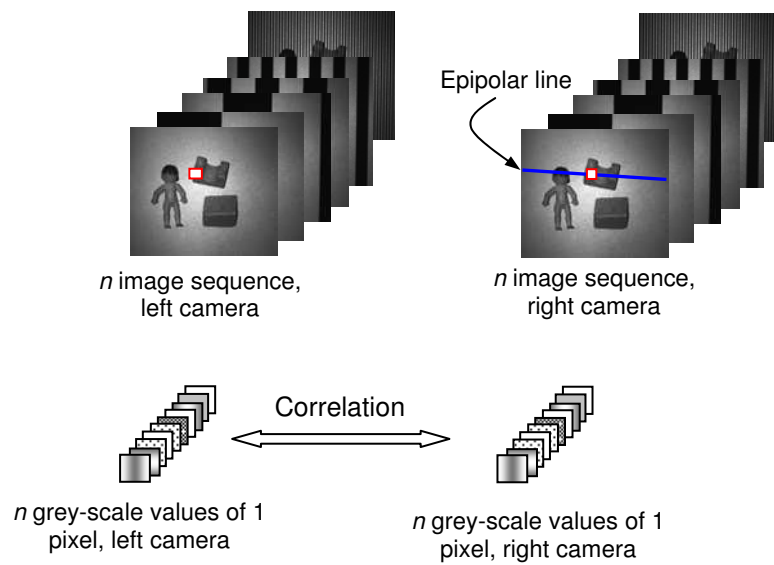


Figure 2.7: Search for the best correlation between the grey-scale values by shifting the pixel of right camera along the epipolar line.

3-D point cloud as well as the 3-D surface reconstruction are shown in Fig. 2.8. In Fig. 2.9, some reference data that will be used in the experimental results is shown. In the left column of Fig. 2.9, the left images from the stereo pairs are shown, Fig. 2.9 (b) illustrates the reference disparity maps that so-called the ground truth disparity maps. the 3-D point cloud and the 3-D surface reconstruction are shown in Fig. 2.9 (c) and Fig. 2.9 (d) respectively.

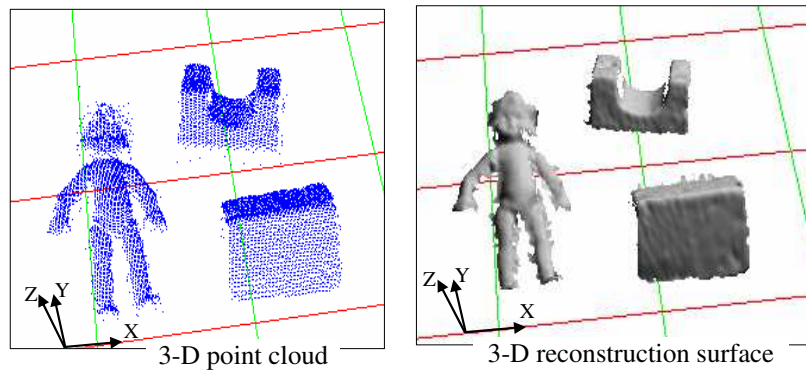


Figure 2.8: Reference data from active stereo vision.

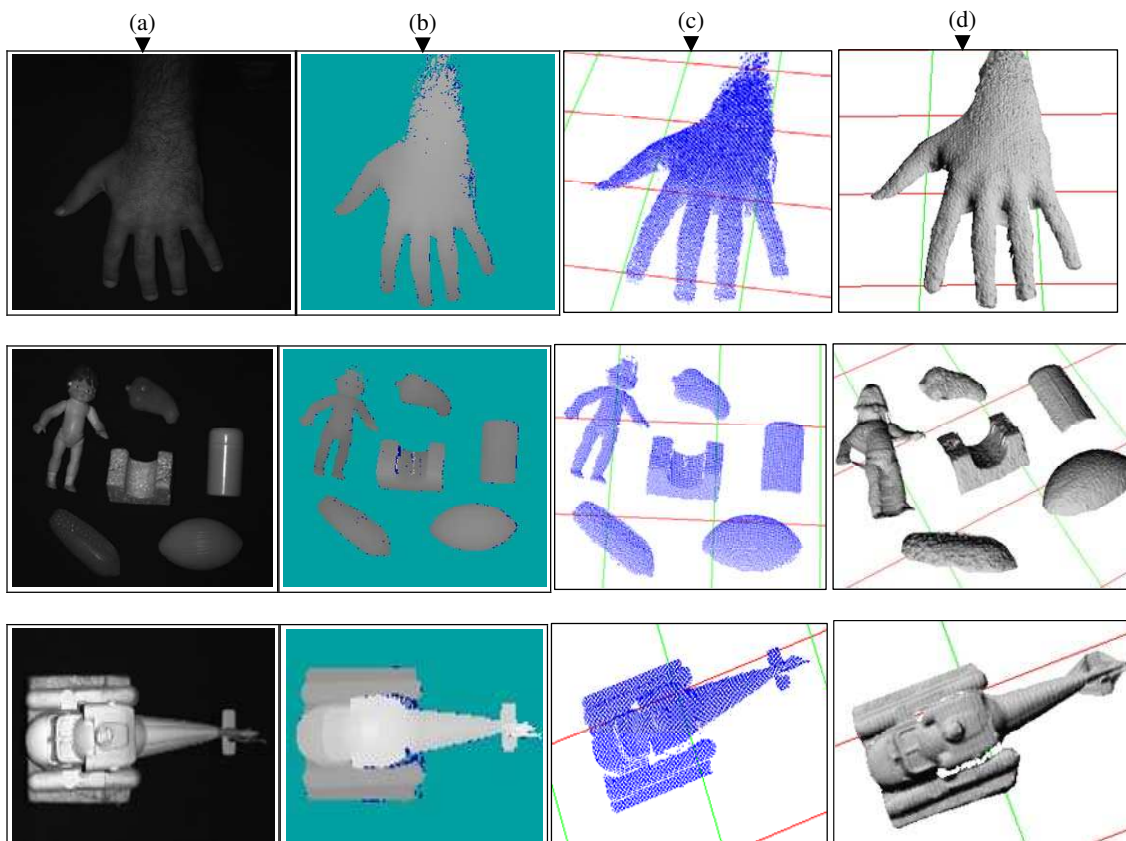


Figure 2.9: Examples of ground truth data used in the experimental results. (a) left image from the stereo pair, (b) the reference disparity map, (c) the 3-D points cloud and (d) the 3-D reconstructed surface.

Chapter 3

Linear Spatial Filters for Disparity Estimation

The frequency-based stereo use phase component, magnitude component, or a combination of them both as the primary matching feature. In this chapter a technique based on using the output of linear spatial filters is presented for characterizing the information present in a vector of filter responses. The left and right images convolving with a set of linear filters tuned to a number of different orientations and scales (using a Gaussian function and its derivatives as a basis functions). To make the correspondence analysis more reliable and robust, the method provides a rich description of the image, with little chance of false matching. The responses of these filters at a given point constitute a vector that characterizes the local structure of the image patch. In the view of [41], an automatic procedure is used to evaluate and optimize the filters set by using a steering theorem [55], [56] and the singular value decomposition (SVD) technique [57]. The correspondence problem can be solved by searching point in the other view where this vector is maximally similar.

As it is known, to simplify the search for correspondence, the image pair is commonly transformed into epipolar geometry; so that the stereo problem is reduced to a one-dimensional search along corresponding scan lines. The offset between x-coordinates in the left and right images is then referred to as a disparity [10] [11]. The foremost contributions of this method arise in detecting and handling the depth discontinuities and the occlusion region to improve the quality of the disparities.

The suggested approach is described by some processing levels (see the concept of the linear spatial filters (LSF) approach in Fig. 3.1), whereby the first level deals with Filter design and optimization. This contains the Gaussian derivatives, filter design, evaluate and optimize filter degree using steering theorem and SVD. The second processing level of the approach is specified by correspondences analysis, which describes the feature extraction, filter response vector (FRV), and estimate the disparity. The improvement of the disparity takes places by the use of adaptive scale filter. Finally,

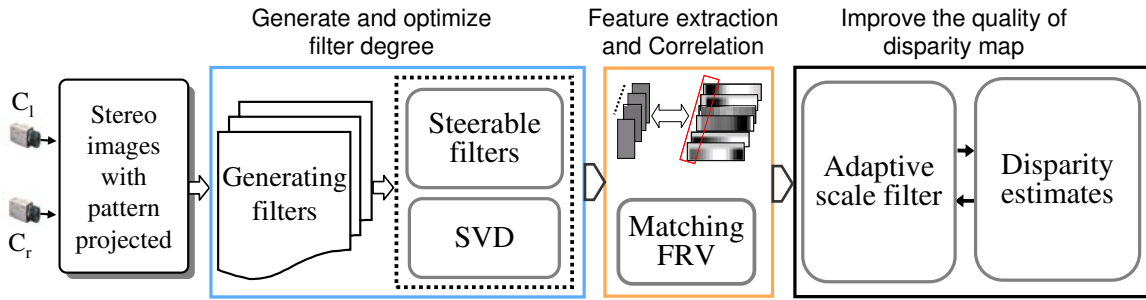


Figure 3.1: Concept of the linear spatial filter (LSF) approach.

the 3-D coordinates of the examined object points are obtained using conventional triangulation.

3.1 Local Analysis of Image Patches

In order to find the correct stereo correspondence, most stereovision algorithms are centered on matching features in one image to the corresponding features in the other. Central to the analysis of these algorithms is to answer the following question: what are the image features to be matched?, How are these features compared to determine corresponding pairs?, and how to handle the errors?.

The accuracy of the results is dependent on the chosen set of filters, so that the implementation and the testing of these ideas requires some particular set of filters to be chosen.

3.1.1 Basis Filter Design: Gaussian Derivatives

Gaussian derivatives provide a basis for series expansion of a local signal [58]. This means that a local image neighborhood can be reconstructed by a linear combination of weighted Gaussian derivative filters. The formula for the n^{th} one dimensional derivative with respect to the dimension x , provided in [59] is:

$$G_n(x) = \frac{d^n}{dx^n} G_0(x) \quad \text{for } n = 1, 2, \dots \quad (3.1)$$

where $G_0(x)$ denoted as the Gaussian basis function

$$G_0(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{x^2}{2\sigma^2}\right)} \quad (3.2)$$

Gaussian derivatives have an explicit scale parameter, σ , and can be generated at any scale. With steerable filters Gaussian derivatives can be oriented in any arbitrary

direction. With automatic scale selection the local scale of a feature can be determined. The object in an image can be normalized by scale which allows recognition under scale changes, the determination of dominant orientation of a neighborhood allows to normalize by orientation.

The filter design carries out by the use of the Gaussian derivative filter bank that is characterized by narrow bandwidths, sharp cutoffs and low overshoots. A key feature of Gaussian filters is that the Fourier transform of a Gaussian is also a Gaussian, so the filter has the same response shape in both the time and frequency domains [59]. The set of filters consisted of rotated copies of filters with impulse responses;

$$F(x, y) = G_n(x) \times G_0(y) \quad (3.3)$$

where, $G_n(x)$ is the n^{th} derivative of a Gaussian and $n = 1, 2, 3$ in our case. The set of basis functions Eq. 3.1 with Eq. 3.2 are referred to here with the initials Gaussian derivative (GD); for example, the GD model for one-dimensional receptive field representations would be a model using Eq. 3.1 and Eq. 3.2.

First derivative of a Gaussian; As well as the first derivative is the well known edge detector, that can be defined as;

$$G_1(x) = K \cdot \left(\frac{-1}{\sigma}\right) \left(\frac{x}{\sigma}\right) e\left(-\frac{x^2}{2\sigma^2}\right); \quad K = \left(\frac{1}{2\pi\sigma^2}\right) \quad (3.4)$$

Second derivative; That is the straightforward extension of the Gaussian first derivative filter described above and can be applied independently in each dimension.

$$G_2(x) = K \cdot \left(\frac{1}{\sigma^2}\right) \left(\frac{x^2}{\sigma^2} - 1\right) e\left(-\frac{x^2}{2\sigma^2}\right) \quad (3.5)$$

Third derivative;

$$G_3(x) = K \cdot \left(\frac{-1}{\sigma^3}\right) \left(\frac{x^3}{\sigma^3} - \frac{3x}{\sigma}\right) e\left(-\frac{x^2}{2\sigma^2}\right) \quad (3.6)$$

Then the general form of a filters set consisting of the n^{th} derivatives of a Gaussian with standard deviation σ at different orientation θ is formulated by

$$F_n^\theta(x, y) = G_n(u) \times G_0(v) \quad (3.7)$$

$$u = x \cos \theta - y \sin \theta \quad v = x \sin \theta + y \cos \theta$$

For a given image $I(x, y)$, its linear (Gaussian) scale-space representation is a family of derived signals $L(x, y; \sigma, \theta)$ defined by the convolution of $I(x, y)$ with the Gaussian kernel, such that

$$L(x, y; \sigma, \theta) = F(x, y; \sigma, \theta) * I(x, y) \quad (3.8)$$

Scale-space σ : The standard deviation of the Gaussian function, or, it is the variance of the Gaussian filter, and for $\sigma = 0$ the resulting filter corresponding to the original image. As σ increases, L is the result of smoothing I with a larger and larger filter, thereby removing more and more of the details which it contains.

The motivation for generating a scale-space representation of a given data set originates from the basic observation that real-world objects are composed of different structures at different scales. This implies that real-world objects, in contrast to idealized mathematical entities such as points or lines, may appear in different ways depending on the scale of observation.

For example, the concept of a "tree" is appropriate at the scale of meters, while concepts such as leaves and molecules are more appropriate at finer scales. For a machine vision system analyzing an unknown scene, there is no way to know a priori what scales are appropriate for describing the interesting structures in the image data. Hence, the only reasonable approach is to consider descriptions at multiple scales in order to be able to capture the unknown scale variations that may occur.

Rotation Parameter θ : For arbitrary rotation of any object in three dimensions, three rotation angles are needed for rotation about the three axes spanning this three-dimensional space. Such rotation parameters are needed in the Gaussian derivative model. To produce such alignment, we rotate the x and y axes so they line up with the intrinsic receptive field axes. These rotation angles are particularly useful for describing the spatial orientation and preferred direction of motion. In the Gaussian derivative spatial model, the rotation parameter for spatial orientation in the $x; y$ plane is given as the angle θ .

The angle θ is the amount of rotation required to make the model receptive field line up its (translated) spatial axes with the real-world spatial axes. The angle θ gives us an estimate of the receptive fields orientation angle in space.

Most frequency-based techniques [4], [41] start processing the image with a family of linear filters tuned at a wide range of orientations and scales of resolution. One then needs to know how many filters are required and how to properly interpolate between the responses. The multiscale - multiorientation image decomposition is then analyzed to detect features and to evaluate the degree to which the chosen filters are independent. In other words, it is important to choose filters that are linear independent to each other and reduce the others which carry no additional information. For this purpose, we use an automatic procedure to evaluate and optimize the chosen filters set, and reduces the overall memory requirement.

3.1.2 Filter Optimization

Oriented filters have found extensive use in many computer vision and image processing task. Such as edge detection, texture analysis, image compression, motion analysis, and 3-D surface reconstruction. In many tasks, it is useful to be able to

tune the orientation of the filters to arbitrary orientation. It is a tedious of these filters. One natural question arises: is it possible to design a set of filters of different orientations and use them as basis functions to synthesize filters with arbitrary orientation?

It has been proved possible to do so with the concept of "steerable filters" introduced in [55]. The term "steerable filters" is used to describe a class of filters in which a filter of arbitrary orientation is synthesized as linear combination of a set of "basis filters".

Consider a set of filters $F(x, y, \sigma, \theta)$, where σ is the scale, $\sigma \in S$, and θ is the orientation, where $\theta \in T$, and x, y are the spatial coordinates. S, T are denoted as the number of scales and orientation. The optimization of the filters set takes place by the following two subsections 3.1.3 and 3.1.4.

3.1.3 Steerable Filters

A function $F(x, y, \sigma, \theta)$ is called steerable if it can be expressed as a linear combination of several rotations of itself. The fundamental idea of steerable filters is to apply distinct "basis filters" that correspond to a fixed set of orientations and interpolate between each discrete response. Thus, one must first decide the number of "basis filters" and corresponding interpolation functions (also known as recombination function).

Let θ_i be the angle of some i^{th} basis filter and $q_r(\theta)$ denote an interpolation functions. As defined in [55], a steering constraint is formulated by

$$F^\theta(x, y) = \sum_{j=1}^M F^{\theta_j}(x, y) q_j(\theta) \quad (3.9)$$

Where M is the number of basis functions required to steer some function $F(x, y)$. Hereafter, it will be more convenient to work in polar coordinates, then F can be expressed as a Fourier series in polar angle

$$F(r, \phi) = \sum_{n=-N}^N a_n(r) e^{in\phi} \quad (3.10)$$

Where $i = \sqrt{-1}$ and N is a discrete length of coefficients. The theorems below were posed by [55] and are included for clarity and completeness of description.

Theorem 1: [55] Steering condition Eq. 3.9 holds for a function F expanded in the form of Eq. 3.10 if and only if the interpolation functions $q_j(\theta)$ are solutions of Eq. 3.11

$$\begin{pmatrix} 1 \\ e^{i\theta} \\ \vdots \\ e^{iN\theta} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{i\theta_1} & e^{i\theta_2} & \cdots & e^{i\theta_M} \\ \vdots & \vdots & \vdots & \vdots \\ e^{iN\theta_1} & e^{iN\theta_2} & \cdots & e^{iN\theta_M} \end{pmatrix} \begin{pmatrix} q_1(\theta) \\ q_2(\theta) \\ \vdots \\ q_M(\theta) \end{pmatrix} \quad (3.11)$$

As well as in polar coordinates

$$F^\theta(r, \phi) = \sum_{r=1}^R g_r(r, \phi) q_j(\theta) \quad (3.12)$$

where $r = \sqrt{x^2 + y^2}$ and $\phi = \arg(x, y)$, $g_r(r, \phi)$ can be any set of function. For any value of n , ($-N \leq n \leq N$), if $a_n(r) = 0$ in Eq. 3.10, then the corresponding n^{th} row of the left hand side and of the matrix of the right hand side of Eq. 3.11 should be removed.

Theorem 2: [55] Let $f(x, y) = W(r)P_N(x, y)$ where $W(r)$ is an arbitrary windowing function and $P_N(x, y)$ is an N^{th} order polynomial in x and y , whose coefficients may depend on r . Linear combinations of $(2N + 1)$ basis functions are sufficient to synthesize $f(x, y)$ rotated to any angle. If $P_N(x, y)$ contains only even (or odd) order terms, then only $(N + 1)$ basis functions are sufficient and Eq. 3.11 can be modified to contain only the even(odd) numbered rows of the left hand side column vector and the right hand side matrix.

In order to design a steerable filter, in addition to design of the basis filters, we also need to know the minimum number of basis filters that are sufficient for steering and the coefficients for the basis filters. As an example, the second derivative of Gaussian function Eq. 3.2 is $G_2(x) = \left(\frac{1}{2\pi\sigma^2}\right) \left(\frac{1}{\sigma^2}\right) \left(\frac{x^2}{\sigma^2} - 1\right) e^{\left(-\frac{x^2}{2\sigma^2}\right)}$. This equation is the product of a second order polynomial $\left(\frac{x^2}{\sigma^2} - 1\right)$ contains only even order and a radially symmetric Gaussian window $e^{\left(-\frac{x^2}{2\sigma^2}\right)}$. Therefore, according to Theorem 2, three basis functions suffice are suffice to synthesize the filters in different orientations.

Using three different orientations, $\theta_1 = 0^\circ$, $\theta_2 = 60^\circ$ and $\theta_3 = 120^\circ$, yields to a set of steerable filters as shown in Fig. 3.3(a) where filter size is 31×31 pixels and the Gaussian sigma is 4. The variance for each filter as measures of variability are shown in Fig. 3.3(b) where each plot is a row-vector containing the variance of each column in the filter (the variance of data set is calculated by taking the arithmetic mean of the squared differences between each value and the mean value). Fig. 3.3(c) shows the output of convolving the filter set shown in Fig. 3.3(a) with the left input image from the stereo pair "Doll image" that are shown in Fig. 3.2.

The other problem is the approximated reconstruction for a given kernel. The basis functions may be predefined or one asks for optimal basis functions for a given kernel.

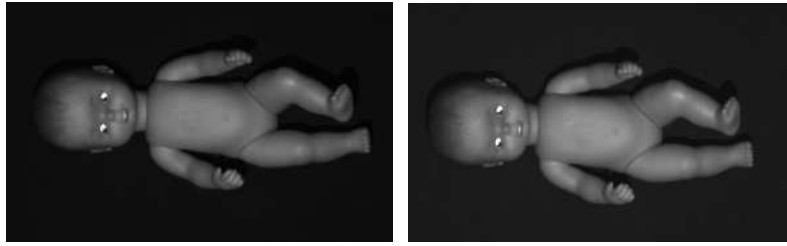
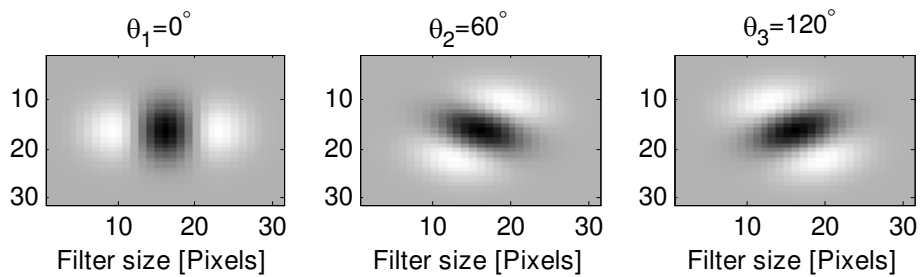
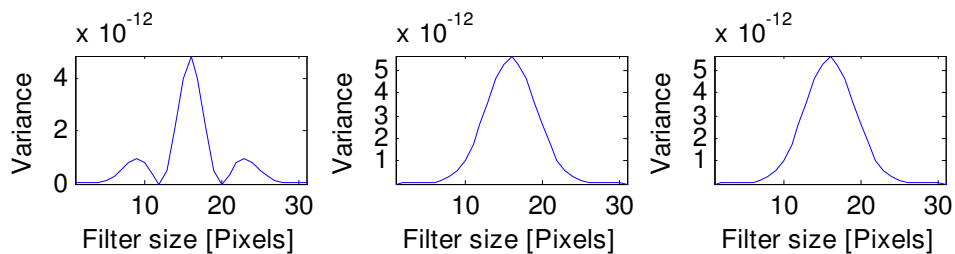


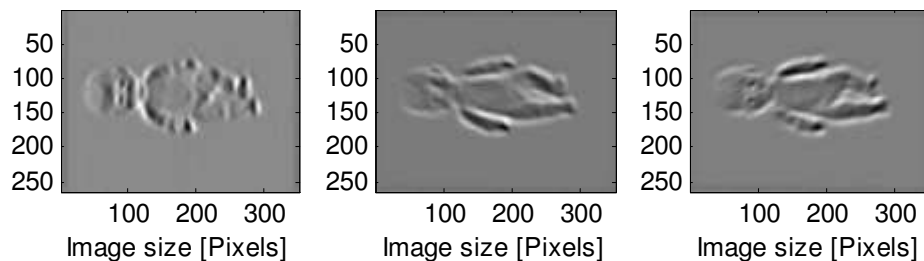
Figure 3.2: Left and right rectified stereo pair images "Doll images" to verify the steerable of Gaussian derivatives filters.



(a) Set of steerable filters for 2nd derivative of Gaussian filter at three different orientations ($\theta = 0^\circ, 60^\circ$ and 120°), the filter scale 31×31 pixels and $\sigma = 4$.



(b) Filter variance. Each plot is a row vector containing the variance of each column in the filter.



(c) The response images. The results of convolving the set of filters shown in (a) with the left input image that is shown in Fig. 3.2.

Figure 3.3: Set of steerable filters. The 2nd derivative of Gaussian filters in different orientations and their responses images.

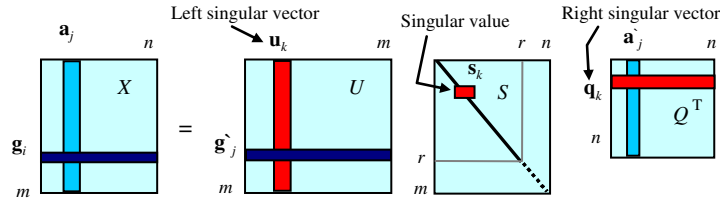


Figure 3.4: Concept of mathematical definition of SVD.

As we mentioned, the implementation and testing of that method requires some particular set of filters to be chosen. Regardless of way a particular set of filters may be chosen, it is useful to use the singular value decomposition (SVD) as an automatic procedure to evaluate the degree to which the chosen filter are independent.

3.1.4 Singular Value Decomposition (SVD)

The singular value decomposition algorithm is one of an effective method to the least squares problems. In the flowing we discourses on the mathematical foundation of the SVD technique [60].

Let X is an $m \times n$ matrix, then it can be alternatively factored as a collection of three matrices, an $m \times m$ orthogonal matrix U (basis function), an $m \times n$ diagonal matrix S with nonnegative diagonal elements in decreasing order, and an $n \times n$ orthogonal matrix Q (recombination function), such that $X = USQ^T$. This decomposition is known as the *singular value decomposition*. Each column u_k of U and q_k of Q represent eigenvectors corresponding to each respective element s_k in the diagonal matrix S , as we can observe from Fig. 3.4.

The diagonal entries of the matrix S are called *singular values* and satisfy $s_1 \geq s_2 \geq \dots \geq s_r \geq 0$. The number of non-zero entries in S is the rank of matrix X (i.e. the number of linear independent column vector in X). If $U = [u_1 u_2 \dots u_m]$ and $Q = [q_1 q_2 \dots q_n]$, then matrix X can be expressed as; $X = \sum_{r=1}^R u_r s_r q_r$. Fig. 3.5 shows the orthonormal basis for the filters in Fig. 3.3 and their responses after using the SVD technique.

In particular, if we are interested in analysis filter banks that represent the rotated version of a prototype kernel, firstly, we exploit the idea of steerable filters. A filter is called steerable if the filter at an arbitrary orientation can be expressed as a linear combination of a set of basis filters, generated from rotations of a single kernel. In our algorithm, we used three basis functions as steerable filters (1^{st} , 2^{nd} and 3^{rd} Gaussian derivatives). The set of filters used consists of rotated copies of filters with impulse responses $F(x, y) = G_n(x) \times G_0(x)$, where $G_n(x)$ is the n^{th} derivative of a gaussian as will explain in the next section. Secondly, The x-y separable steerable approximations of filter kernels were generated by a singular value decomposition. Singular values

and corresponding singular vectors contain complete information about the filter.

3.1.5 Gaussian Filter Bank

Let a spatial filter with finite impulse response be represented as a one-column vector, F_i , by writing out its entries row by row. Its dimension is $n \times 1$, where n is the number of pixels in the support of the filter. If an image patch, I , (of the same size and shape as the support of the filter) is also represented as an $n \times 1$ column vector, then the result of convolving the image patch by the filter is simply the inner product of these two vectors, as we can show in Fig.3.6. Taken together, a set of spatial filters forms a matrix F . This is a convenient representation of the linear transformation that maps image patches to a vector of filter responses. For an image patch represented as a vector I , the filter response vector is simply $v = F^T I$.

Applying the singular value decomposition on the set of spatial filters F^T yields $F^T = USQ^T$. The number of non-zero entries in S , is the rank r , or the dimension of the vector space spanned by the filters. The first r columns of Q form an orthonormal basis set for this vector space, ranked in order of the visual patterns to which this particular set of filters is most sensitive. The corresponding singular values indicate how sensitive.

Our filter set used in that method consists of rotated copies of filters with impulse responses $F(x, y) = G_n(x) \times G_0(y)$ where $n = 1, 2, 3$. The scale σ was chosen to be the same in both the x and y directions. Filters at five scales were used, with the area of the filters increasing by a factor of two at each scale, and the scale $\sigma = \frac{w}{8}$ where $w \times w$ is the filter size with $w \in \{5, 7, 10, 14, 20\}$. Nine filters (nine orientation) at five scales would give 45 filters, a more efficient approach is to apply a few filters corresponding to a few angles and interpolate between the responses. As an example of this decomposition, the orthonormal basis for the set of filters (at the scale $\sigma = 4$ and filter window size 29×29) is shown in Fig. 3.8. Table 3.1 shows the pseudo code for designing the set of filters.

From the Gaussian derivatives which presented in section 3.1.1, one can say that, the set of filters are steerable, wherever, the first derivative of a Gaussian, Eq. 3.4, is the product of a first order polynomial (only odd terms) and a radial symmetric Gaussian window, according to theorem 2, two basis functions suffice to synthesize G_1^θ . The second derivative of a Gaussian, Eq. 3.5, is also a product of a second order polynomial (only even order terms) and a radial symmetric Gaussian window. So, by theorem 2, three basis functions are sufficed. Similarly, the third derivative, Eq. 3.6, required four basis functions to synthesize G_3^θ . For Gaussian derivatives in particular, it turns out $n + 1$ different orientations are required for n^{th} Gaussian derivative [55].

Experimental test: Consider, a set of four steerable filters consisting of the first derivative of Gaussian at four different orientations 0° , 30° , 60° and 90° , $G_1^{0^\circ}$ is the

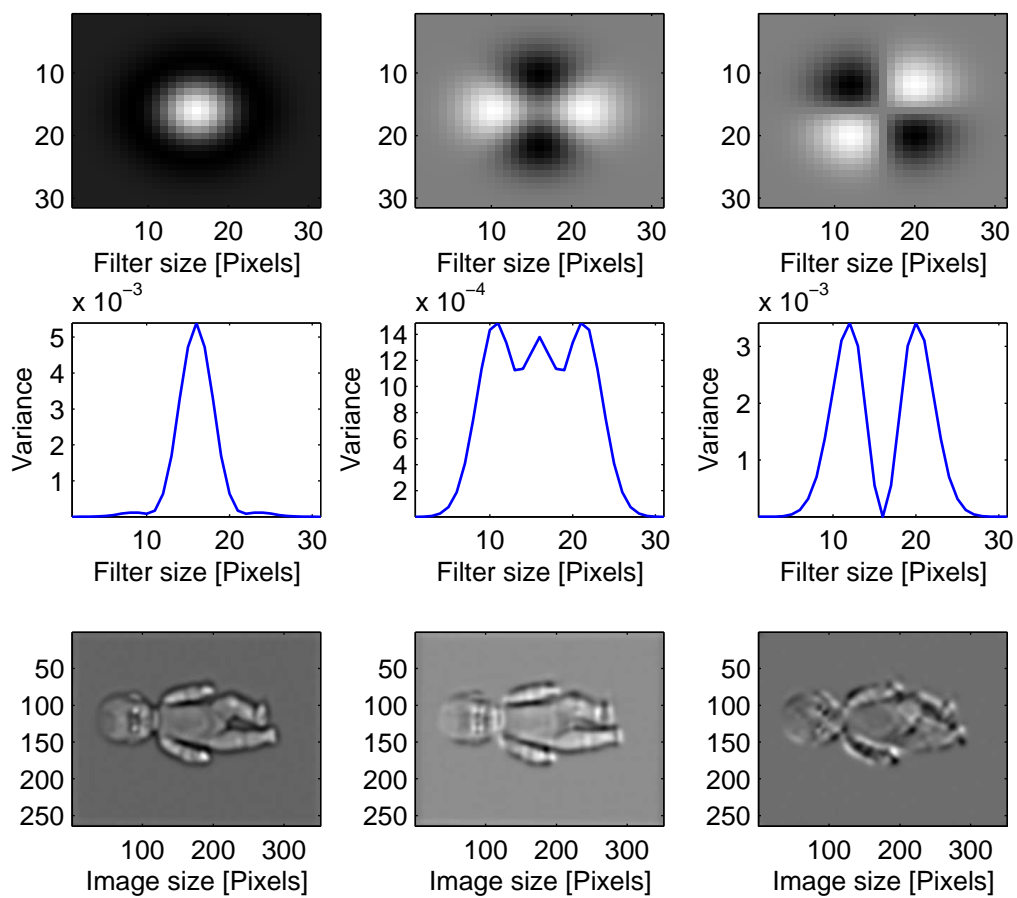


Figure 3.5: (Top) Orthonormal basis for the filters set shown in Fig. 3.3 as the output of SVD. (Middle) Plots of row-vector containing the variance of each column in the filter where x-axis is the filter length and the y-axis is the variance values. (Bottom) The filter response images of the corresponds filters.

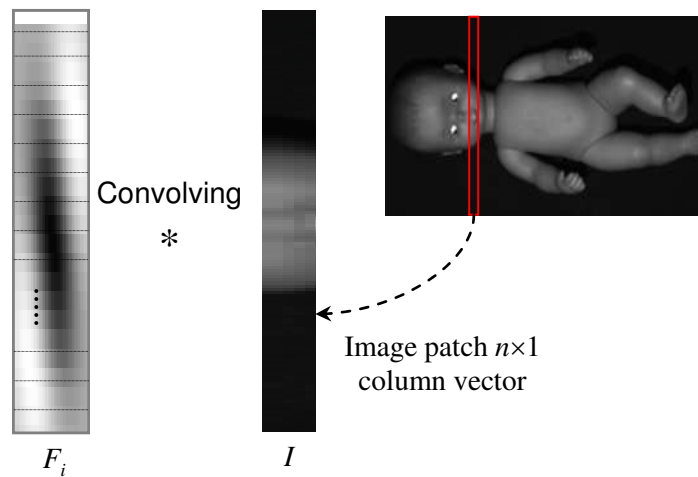


Figure 3.6: Convolve a one column vector spatial filter, F_i , with a one column image patch, I .

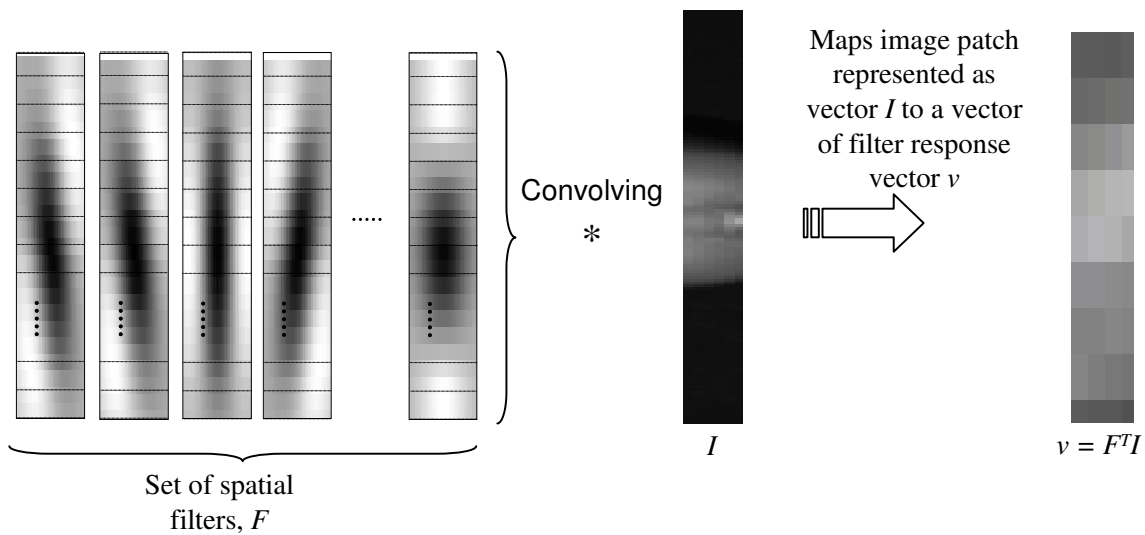


Figure 3.7: A linear transformation maps image patches I (where I is image patch $n \times 1$ column vector from the input image "Doll Image") to a vector of filter responses, $v = F^T I$.

Table 3.1: Pseudo code for design a set of filters used

First Derivative:			
1:	$Filter = G_0^T \cdot G_1$		
2:	$Filter = G_0^T \cdot G_1$	Rotated	90°
Second Derivative:			
3:	$Filter = G_0^T \cdot G_2$		
4:	$Filter = G_0^T \cdot G_2$	Rotated	45°
5:	$Filter = G_0^T \cdot G_2$	Rotated	120°
Third Derivative:			
6:	$Filter = G_0^T \cdot G_3$		
7:	$Filter = G_0^T \cdot G_3$	Rotated	45°
8:	$Filter = G_0^T \cdot G_3$	Rotated	90°
9:	$Filter = G_0^T \cdot G_3$	Rotated	120°

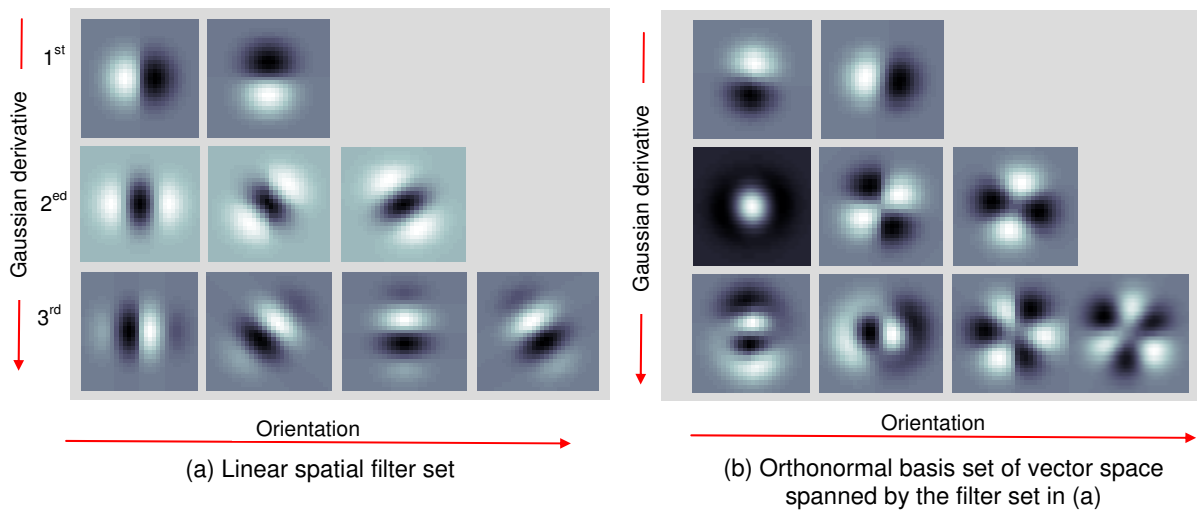


Figure 3.8: List of suggested linear spatial filters according to the filter set design as shown in Tab. 3.1, successive derivatives and rotations of Gaussian curves.

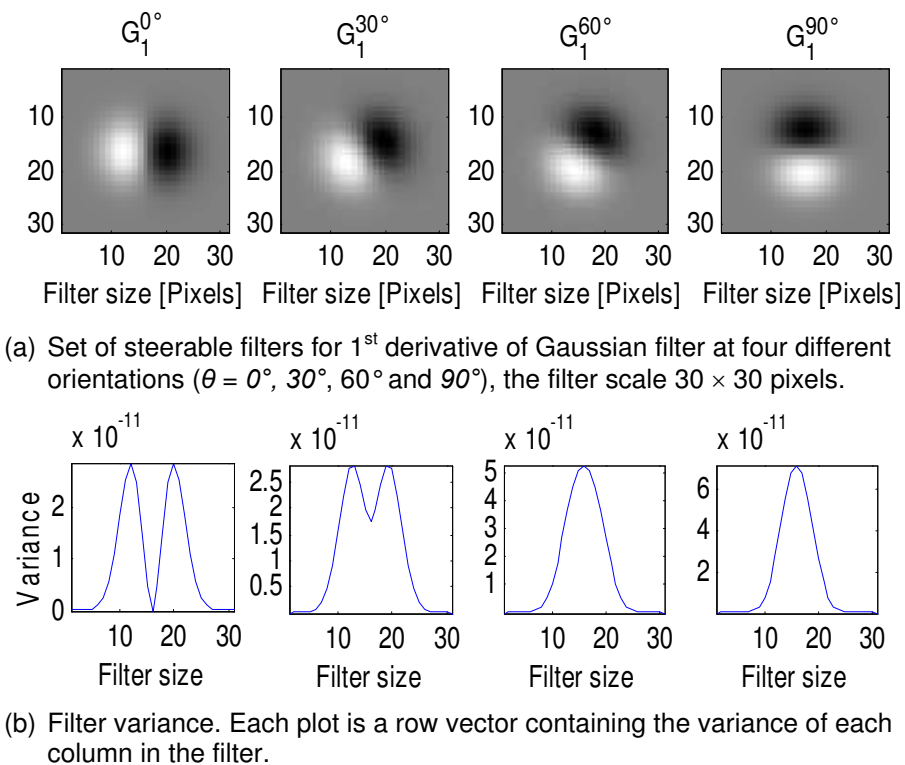


Figure 3.9: Set of filters. Four filters of the 1st Gaussian derivative at four different orientations $0^\circ, 30^\circ, 60^\circ$ and $90^\circ, \sigma = 4$ and filter size 31×31 pixels.

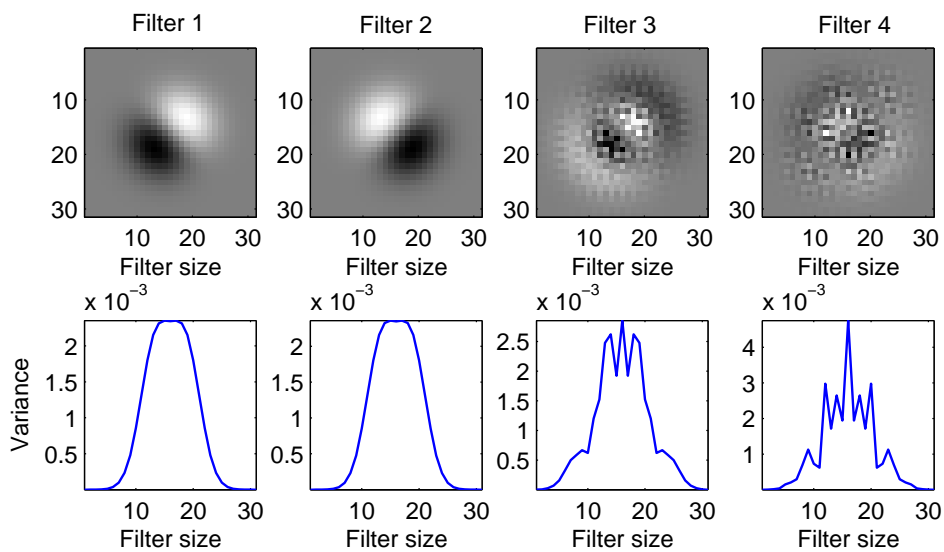


Figure 3.10: The adaptive filter set for the chosen filter set that are shown in Fig. 3.9.

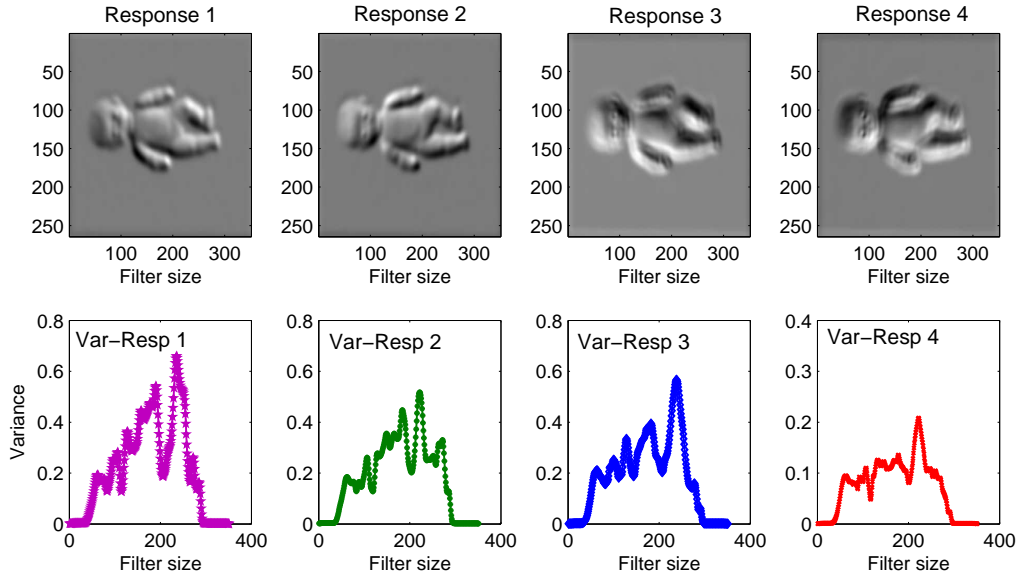


Figure 3.11: Filter response images. (Top) is the filter response images obtained from convolving the adaptive set filters that are shown in Fig. 3.10 with the "Doll image" that is shown in Fig. 3.2. (Bottom) is a row vector of variances the filter response images.

first derivative with respect to (horizontal) of Gaussian, $G_1^{30^\circ}$ is the second filter which is $G_1^{0^\circ}$ rotated by 30° , $G_1^{60^\circ}$ and $G_1^{90^\circ}$ are the third and fourth filter. The filter size $w = 31 \times 31$ pixels and $\sigma = 4$ as we can see in Fig. 3.9 (a). The variance as indicator for the variability in each filter are shown in fig. 3.9 (b). By using the singular value decomposition as an automatic procedure to evaluate the degree to which the chosen filter are independent. The vector space spanned by these four filters is only two-dimensional. That means only two filters are needed, since the other may be the weighted sum of these, and thus carry no additional information.

Fig.3.10 shows the adaptive filter set for the chosen four filters that are shown in Fig. 3.9. Fig.3.10(Top row) illustrates the reconstructed filters (Filter 1, Filter 2, Filter 3 and Filter 4) where we can ignore "Filter 3" and "Filter 4" which are carry no additional information, as we will verify that in the following. The variance of each filter is computed and plotted in Fig. 3.10(Bottom) Where the curves indicate that, the latest two filters (Filter 3 and Filter 4) looks like noise (carry no additional information). The degree of how sensitive the filter is indicated from the corresponding singular values. For these four filter set, the corresponding singular values are 0.00020854 ; 0.00013189 ; 8.5014×10^{-7} and 5.3693×10^{-7} , respectively. One can easily observe that, the latest two filters have singular values close to zero. Therefore, they are redundant and so were discarded. By looking at the filter response images, Fig.

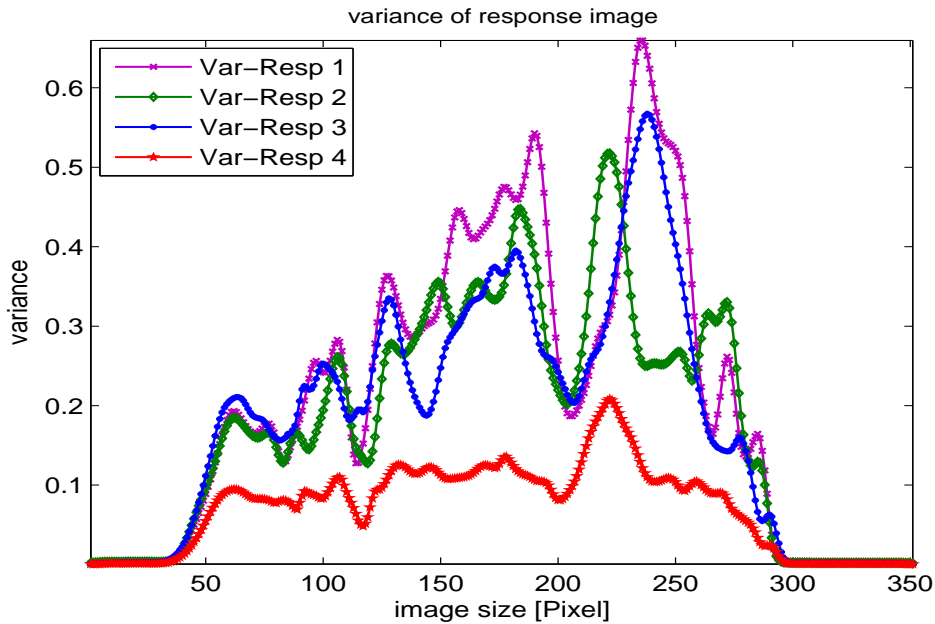


Figure 3.12: The variance of each response image that reflects the amounts of information in each response image.

3.11(Top) shows the filter response images obtained from convolving the input image "Doll image" with the adaptive filter set that are shown in Fig. 3.10. The row vector containing the variance of each column from the filter response image is shown in Fig. 3.11 (Bottom). Fig. 3.12 shows all the same curves of variances each filter-response-image ($Var - Resp1$, $Var - Resp2$, $Var - Resp3$, $Var - Resp4$) where the area under the curve characterize the amount of information in the response image, so it can observe that the "Response 3" and "Response 4" carry only redundant information therefore they can be ignored.

3.2 Matching Filter Responses

The proposed method is based on the output of convolving the left and right image with a bank of linear spatial filters at different number of orientations and scales. In this section we will explore how should filter response vectors be compared. Although corresponding filter response vectors in the two views should be very similar, differences in shading and foreshortening mean that they will rarely be identical. A variety of similarity measures can be used to compare two vectors, including the angle between them or some norm of their vector difference. These measures are zero when the filter response vectors are identical and otherwise their magnitude is proportional to some aspect of the difference between potentially corresponding image patches. It

turns out that any number of such measures do indistinguishably well at identifying corresponding points in a pair of stereo images, except at depth discontinuities. Near an object boundary, the large spatial filters lie across an image patch containing the projection of more than one surface. Because these surfaces lie at different depths, and thus have different horizontal disparities, the filter responses can differ considerably in the two views, even when they are centered on points that correspond. This difference is much more appropriately treated as an outlier, making least-squares approaches inapplicable. We used the sum of absolute differences of corresponding filter responses, which are less sensitive under these conditions.

The sum of absolute differences is calculated by computing the correlation value using

$$E(x, y) = \sum_k |F_k * I_l(x, y) - K_k * I_r(x + disp, y)| \quad (3.13)$$

where $disp$ is the candidate disparity, and $disp \in d$ where d is the range of disparity determined by a priori estimates of the range of horizontal disparities. F_k is the current filter at scale k that ranges from the scale interval. I_r and I_l are the right and left image respectively. The symbol $*$ denotes convolution operator. For each candidate corresponding point, the difference in filter response vectors, E , is computed. The positional offset to the point that minimizes this difference is recorded as the positional disparity at pixel $P(x, y)$ in the left view. This procedure is repeated for each pixel in both images, providing the disparity maps for both the left and right views.

This procedure of using filter outputs for matching is depicted in Fig. 3.13. The outputs of the convolution are presented in the n^{th} left and right responses images, as shown in Fig. 3.13(a). The responses for a given pixel $P(x, y)$ is usually formed by the sequence of response values for that pixel across the outputs of the filter set, FRV, shown in Fig. 3.13(b).

Because the disparities are offsets in terms of image coordinates, the disparity values for corresponding points in the left and right images should have equal magnitudes, but opposite signs. Whenever the support of the filter set lies almost entirely on a single surface, the disparity estimates are correct. Even close to depth discontinuities, the recovered disparity is quite accurate, despite the responses from the larger filters being contaminated by lying across surfaces at different depths.

The recovered disparity map for stereo pair flat surface is presented in Fig. 3.14, disparity values at each image location are presented as grey zero horizontal disparity, and brighter or darker shades for positive or negative disparities. This is shown in Fig. 3.14(middle), which illustrate the left and right recovered disparity maps also a reference disparity map from left to right. The recovered disparity values clearly matches the surface quite well as compared with the reference disparity map. The 3-D point cloud and the 3-D surface reconstruction of the examined object, shown on bottom, provide quite good results. That is because the support of the filter set

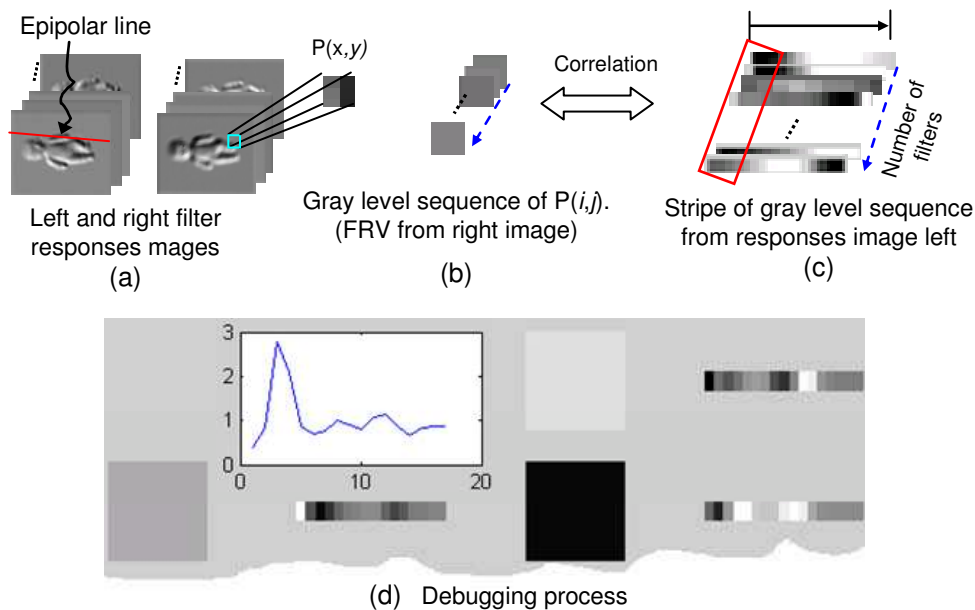


Figure 3.13: Matching the filter outputs by shifting the pixel of responses left image along the epipolar line.

lies almost entirely on a single surface. Another example is the stereo pair of "a hand" shown in Fig. 3.15, the computed disparity map shown on the top-left of the figure shows some error pixels in black or white colors. In such region there is no corresponding point in the other view and the recovered disparity appear as noise or as an error. These disparity maps can be improved using the adaptive scale filter selection.

3.3 Depth Discontinuities and Occluded Region

In a stereo pair of images, it is expected that in the output of a set of filters at a range of orientations and scales, the corresponding image patches should be quite similar. This expectation is reasonable when all of the spatial filters are applied to image patches which are the projection of a single surface. But, when larger spatial filters straddle depth discontinuities possibly including occluded regions, the response of filters centered on corresponding image points may differ quite significantly.

In the scene that contains more than one level surface, which has a vertically oriented edge; there are often regions visible to one camera, but not visible to the other. In this region there is no corresponding point in the other view and recovered disparity estimates appear as noise (error). This situation is depicted in Fig. 3.16(a), as shown in the figure, the far surface of the region D which is visible only to the right camera, Similarly, region B visible only to the left camera.

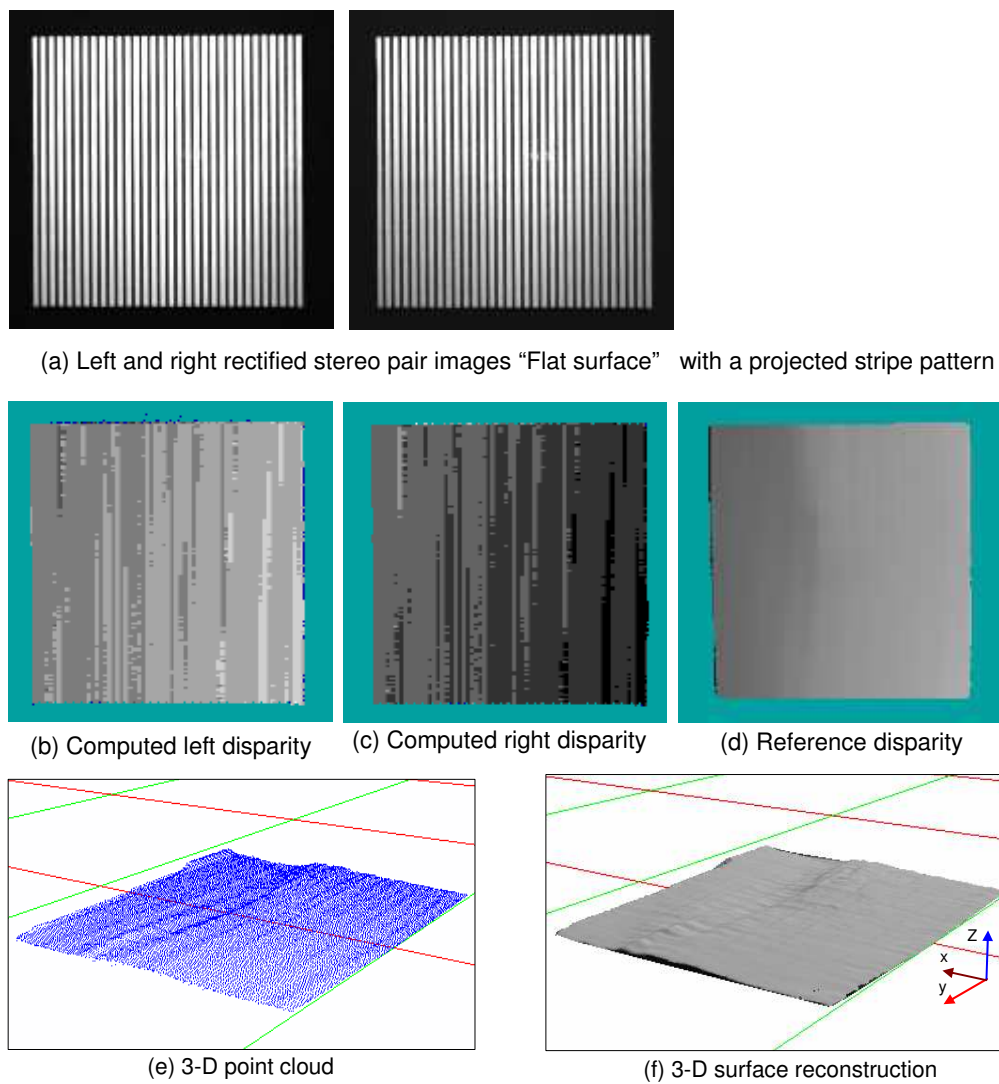


Figure 3.14: Surface reconstruction for a flat surface. (a) Rectified stereo pair images with stripe pattern projector, (b) left computed disparity maps, (c) right computed disparity map, (d) the ground truth disparity (Reference disparity), (e) the 3-D point clouds, (f) the 3-D surface reconstruction.

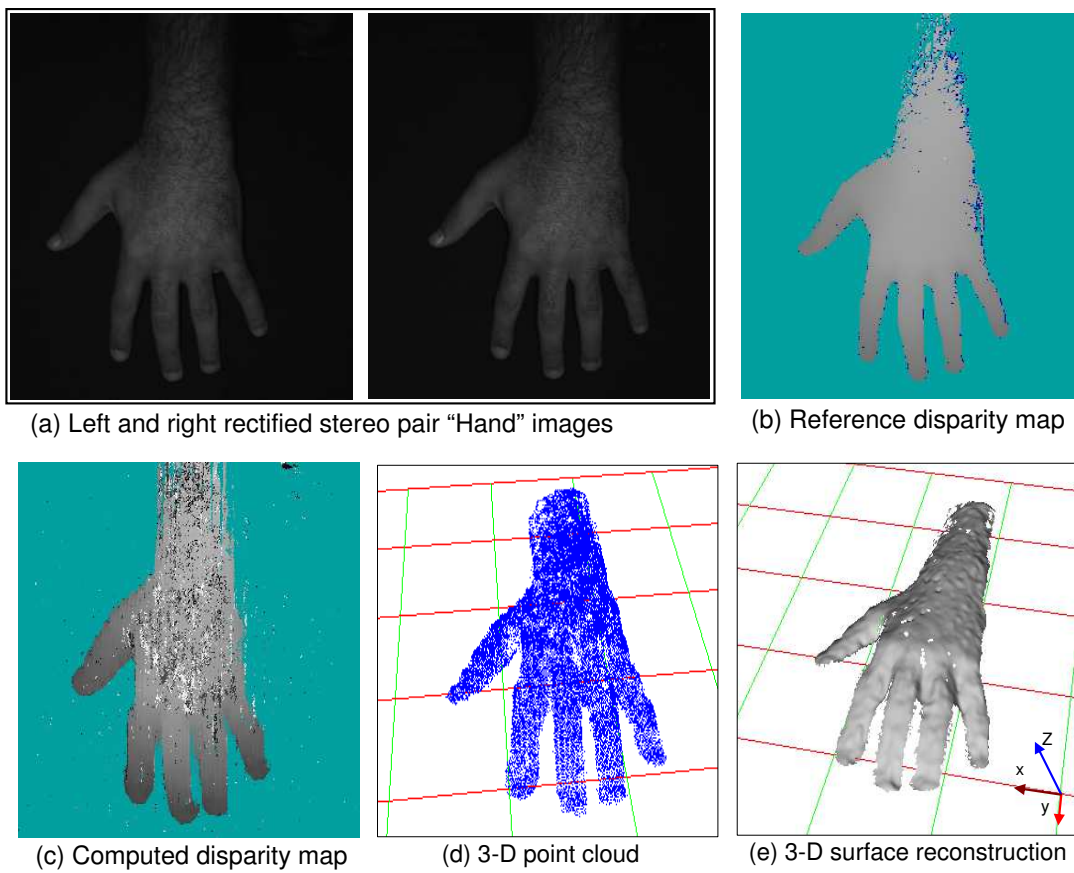


Figure 3.15: (a) The left and right images of stereo pair "Hand", (b) the reference disparity map, (c) the computed disparity map, (d) the 3-D point cloud, and (e) the 3-D surface reconstruction.

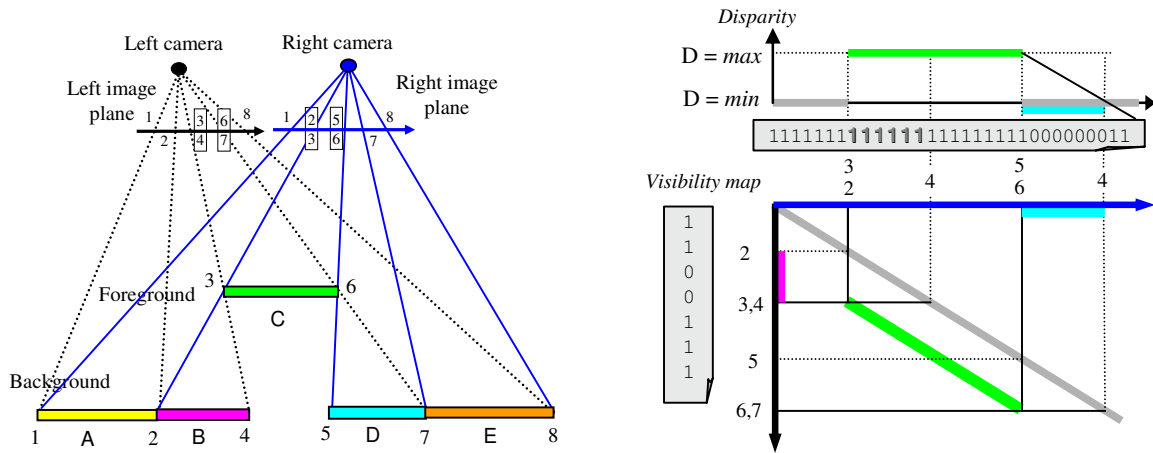


Figure 3.16: Stereo geometry and visible projection regions. (Left) Stereo geometry with parallel axes for depth discontinuities view. (Right) Visibility map. The zeros mark the regions detected to be visible only from one of the two views.

The identification of the occluded regions is important for several reasons. First, nonsense disparity estimates within occluded regions can be eliminated entirely from the recovery of viewing parameters. Second, the knowledge of which regions are visible in the two views and which are not can be used when refining disparity estimates, especially to prevent nonsense disparity estimates within occluded regions from being propagated to neighboring visible regions from the two views. Finally, identified occluded regions make certain aspects of the three dimensional structures of a scene explicit, including the delineation of the occluding contours.

The main idea for detecting and localizing the occluded regions arises in the fact that, the occluded regions in one image lie in the disparity map of the other image. This suggests that the best cure for finding occluded regions in one image lies in the disparity estimates for the other image. Since the occluded regions in one image include exactly those points for which there is no corresponding point in the other image. For detecting and localizing the occluded regions in one view, we define a binary *visibility map*, $V(x, y)$ for one view, as 1 at each image position that is visible in the other view, and 0 otherwise (i.e., occluded region). The disparity values for each point in the left image are signed offsets that give the coordinates of the corresponding point in the right image, if the visibility map for the right image is initially all zero, it can be filled in systematically as follows:

For each position in the left image, its disparity is in the range $[d_{min}, d_{max}]$, set the corresponding position in the right visibility map to 1. The other positions in the visibility map remain 0 that have no corresponding position in the right image and are considered occluded.

The description of this manner is shown in Fig. 3.16 and Fig. 3.17, by locking

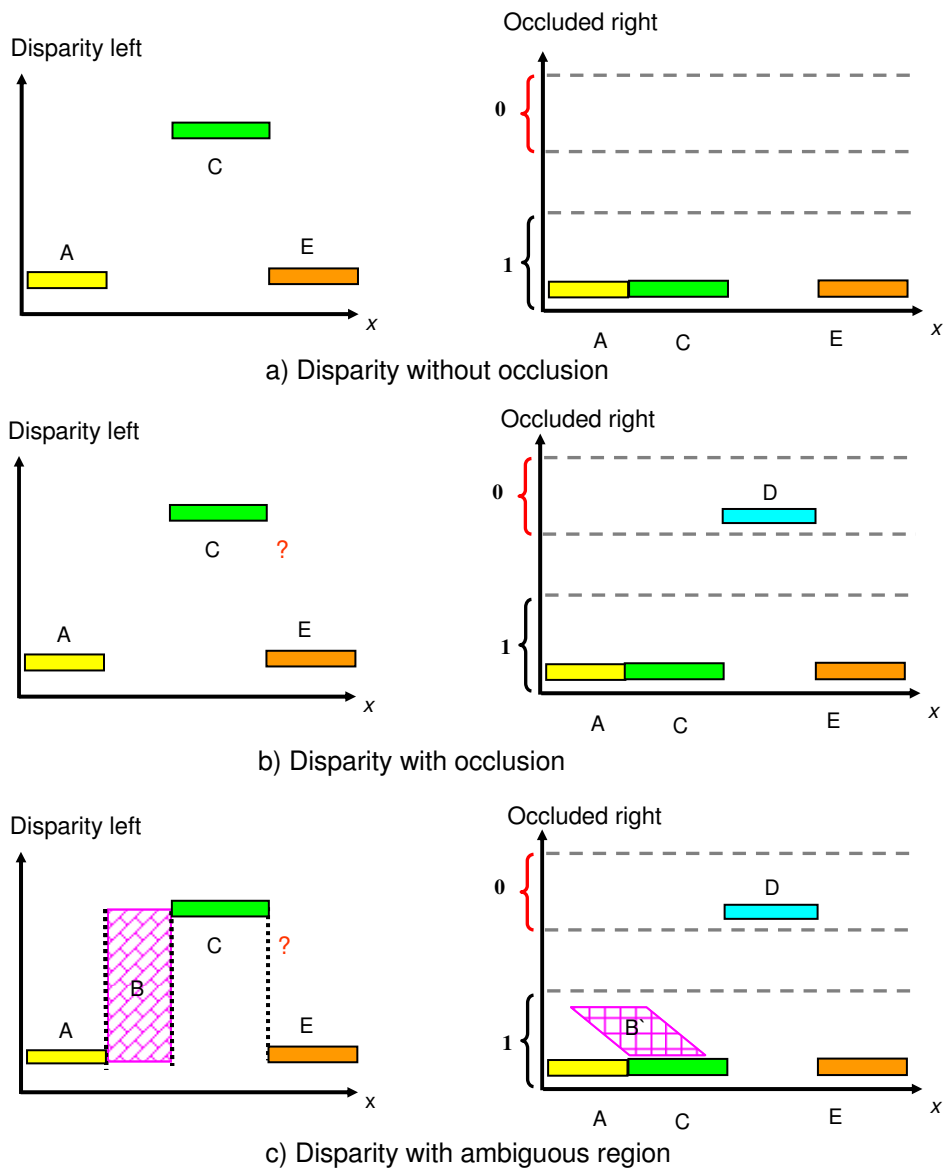


Figure 3.17: Occlusion interpretation. The disparity map of the left image (left column), detecting occlusion map of the right image (right column).

for the regions A , C and E in Fig. 3.16(a) and Fig. 3.17(a) it can be observed that all pixels on both views are visible and their disparity values tend to be unambiguous. Therefore the visibility map signed to 1 at these regions. The other situation comes into view region D , it is visible only to the left view, therefore, there are no corresponding points in the right view. So this region is quite likely occluded and the visibility map signed to 0, see Fig. 3.17(b). In addition, it can be observed that disparities in occlusion regions are usually ambiguous e.g., the region B , which is not visible in the right view.

An example of a visibility map computed in this manner is shown in Fig.3.18. The pair of stereo "Tsukuba images" is shown the top. The white area in the left and right visible maps mark the regions that contains no correct disparity value that can be assigned because there is no corresponding point in the other view.

3.3.1 Adaptive Scale Selection

In this section we discuss a correct treatment for coping with the initial difficulties due to the depth discontinuity, using adaptive scale selection. Whenever a substantial area of a filter is applied to a region of significant depth variation, such as inappropriately large scale filter should be selectively ignored.

From an initial disparity map, it is possible to estimate where such inappropriately large-scale filters are being used by applying the following procedure.

At each position in the image, the median disparity is determined over a neighborhood equal to the support of the largest spatial filter used for stereo matching. Over this same neighborhood, the difference between each disparity estimate and this median disparity is determined. These differences are weighted by a Gaussian at the same scale as the filter, since the center of the image patch has a greater effect on the filter response. The sum of these weighted disparity differences provide a measure of the amount of depth variation across the image patch affecting the response of this spatial filter. When this sum exceeds an appropriately chosen threshold, it may be concluded that the filter is too large for its response to be useful in computing correspondence. Otherwise, continuing to make use of the outputs of large spatial filters provides stability in the presence of noise.

Taking the difference between the disparity estimate at each location and median disparity value of the neighborhood makes the implicit assumption that surface are nearly fronto- parallel. Instead, it will generally be better to use the difference from the best-fitting plane of disparity over the neighborhood.

To record the results, the notation of a *scale map* is introduced. At each position in an image, the scale map, $S(x, y)$, records the scale of the largest filter used in computing stereo correspondence. For the computation of initial disparity estimates, all the scales of spatial filters are used. From initial disparity estimates, it is possible to modify the scale map. For each position in the scale map, if it is determined that

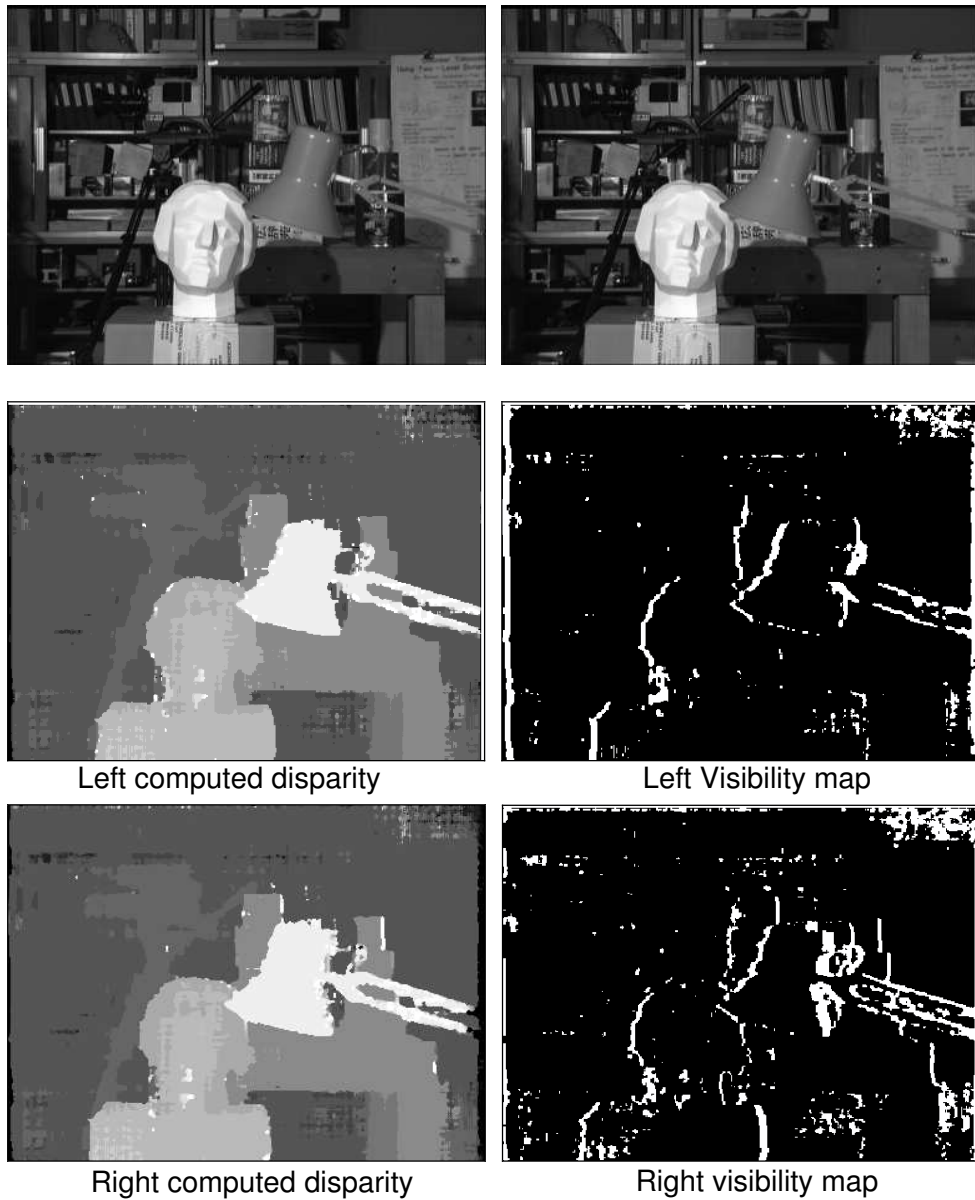


Figure 3.18: Visibility map. The white area in the left and right visible maps mark the regions that contain no correct disparity value.

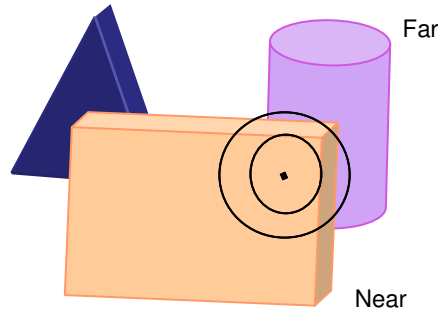


Figure 3.19: Scale selection in the depth discontinuities scene.

an inappropriately large-scale filter was used, then the scale value at that position will be decremented. Otherwise, continue to the next position.

Once initial estimate disparity was obtained, an additional information becomes available which can be used to improve the quality of the disparity estimates. This additional information includes estimates of viewing parameters (camera parameters), location of occluded regions and the appropriate scale of filters to be used for matching. Furthermore the 3-D surface reconstruction can also be obtained using conventional triangulation to depict what the scene might look like from a new viewpoint.

3.4 Experimental Results

An implementation of this approach using the outputs of a number of spatial filters at a variety of orientations and scales as the basis for establishing correspondence have proven to give quite good results.

The algorithm has been implemented and tested on a variety of stereo images with structured light pattern (stripe pattern). Fig. 3.20 and Fig. 3.21 show a scene that contains a "Cylinder" object. Fig. 3.20 shows a set of filter response images which are the outputs of convolve a set of linear spatial Gaussian derivative filter (nine filters) at different orientations (1^{st} derivative at 0° and 90° , 2^{nd} derivative at 0° , 45° and 120° , and the 3^{rd} derivative at 0° , 45° , 90° , and 120°) with the left image. The responses of these filters at a given point constitute a vector that characterize the local structure of the image patch. The filter response vectors in the two views are correlated. This correlation is repeated for each pixel in both images providing the disparity map. Fig. 3.21 shows the initial disparity map which computed by our suggested algorithm. That initial disparity map will be improved by adapting the scale filter selection procedure. By comparing the initial disparity map Fig. 3.21(b) with the ground truth disparity map (reference disparity) Fig. 3.21(a). We can see some outlier especially in the boundary of the cylinder, Fig. 3.21(c) shows a scan line

Table 3.2: Pseudo code for our LSF Stereo Matching Algorithm

Given: A pair of gray scale images, disparity range.
 Let $I_l(x, y)$ and $I_r(x, y)$ are the left and right images.
 Let $disp$ be the candidate disparity in the allowable disparity range d .
for each pixel $P(x, y)$ in I_l
 for each candidate disparity $disp$
 Compute the error measure $E(x, y, disp)$

$$= \sum_k |F_k * I_l(x, y) - K_k * I_r(x + disp, y)|$$

 Let $\Delta(x_l, y_l)$ be the choice value which yields the min value of E
 Compute the visibility map $B(x, y)$ and the scale map $S(x, y)$
 if $\Delta(x_l, y_l) > 0$
 $B(i, j) = 1$
 else
 $B(i, j) = 1$, *there is no disparity value therefore the point is occluded.*
 end
end
end

profile taken from the computed disparity (blue line) and the true disparity (red line). The regions that show the error in the computed disparity are marked by ellipsoids on Fig. 3.21(c).

In practice, from the initial computed disparity map and with a suitable numbers of iteration as depicted in in Fig. 3.21 we can follow the improvements in the quality of the computed disparity. For each point in the scale map that is contain all scale filters, if an inappropriately large-scale filter was used, then the scale value at that position will be decremented. I.e. whenever a substantial area of a filter is applied to a region of significant depth variation, such as inappropriately large-scale filter should be selectively ignored. A final result after ten iterations is shown in Fig. 3.23, from that graph we can say that the proposed method has a superior performance in comparison with a traditional stereo algorithms (area-based algorithm).

The outlier of the area-based algorithm arise due to the correlation depending on the intensity profiles of the two images, they have a very good matching accuracy in the continuously smooth surface areas. But a large deformation of matching results from projective distortion and smoothing occurs in vicinity of depth discontinuity. Also, use of the same-size matching window in both images means that objects are assumed to be basically planar and viewed from the front.

The final disparity maps for other experimental scenes are shown in Fig. 3.24. One can compare the computed disparity map in the middle-right with the true map in the middle-left. The color error maps, in the bottom-left, show the outlier pixels in (magenta), and the invalid pixels in the computed disparity map appears as (blue) pixels, where correct regions are excluded and shown in gray level from 32 to 64 color

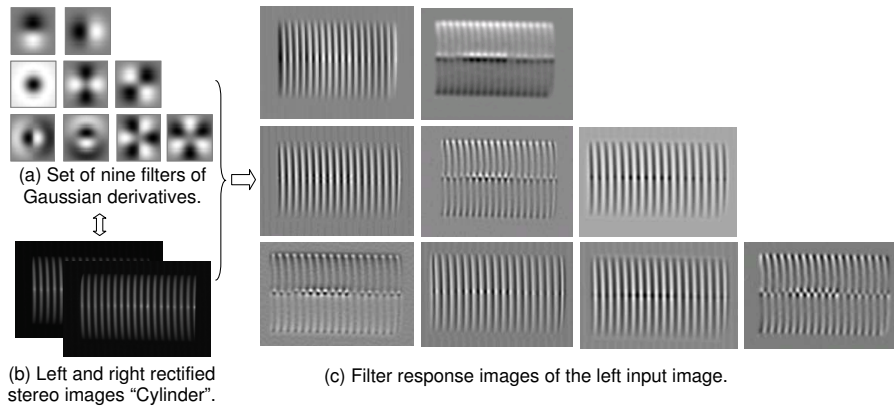


Figure 3.20: Convoluting a left and right images with a set of linear spatial filters. (a) The set of nine Gaussian filters, at filter size 31×31 . (b) Stereo pair images of "Cylinder" object. (c) The response images of the left image.

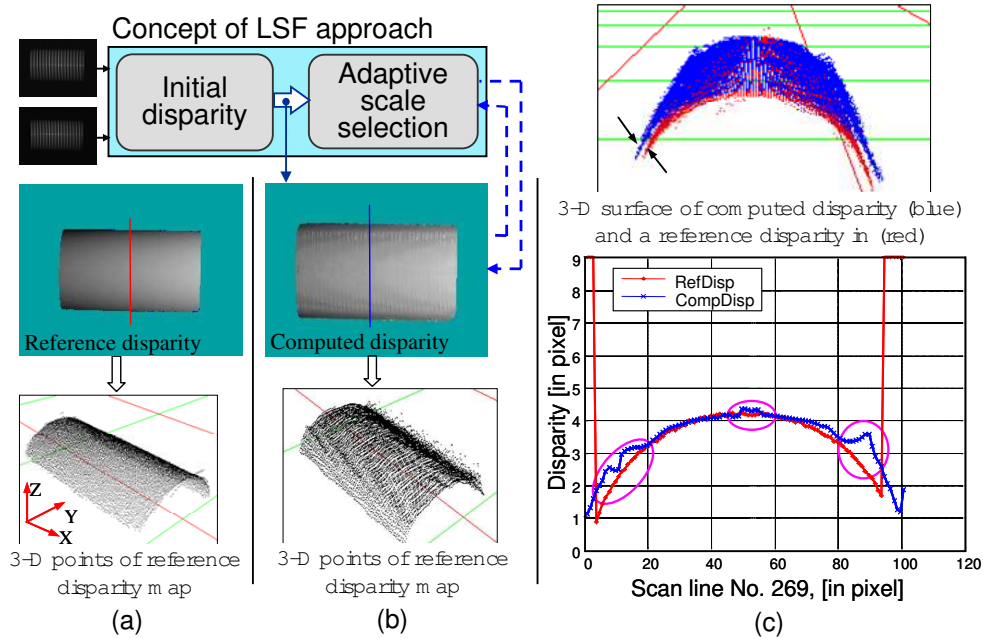


Figure 3.21: Initial estimated disparity map for a stereo pair image of a "Cylinder" object computed from the suggested linear spatial filter (LSF) approach.

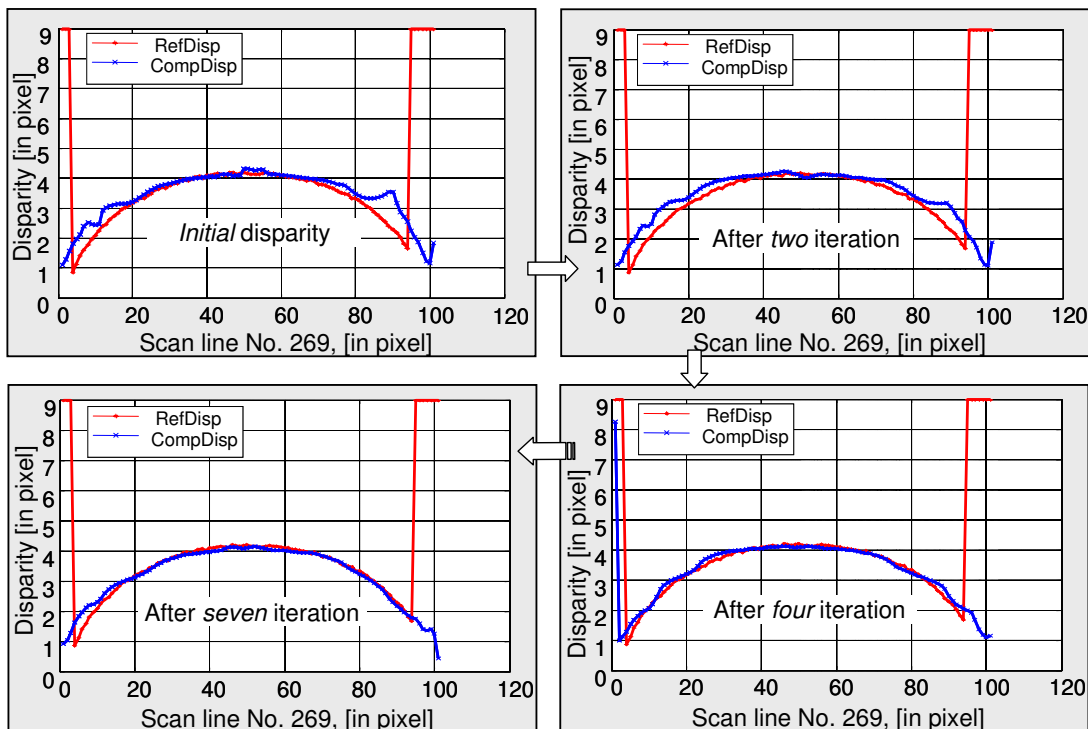


Figure 3.22: Progress the iteration of our algorithm to follow the results develop. The scan line from the computed disparity shown in the blue line and the reference scan line shown the red line in each iteration.

map. Finally, the 3-Dimension points well appear in bottom-right.

Fig. 3.25 shows the final results provided from our linear spatial filter algorithm after ten iterations for the "Tsukuba" stereo images from that are shown in Fig. 3.18. We can compare that final results with that presented in Fig. 3.18 to observe some improve for the initial computed disparity, therefore the courant results in Fig. 3.25 are quite accurate than the initial computed results.

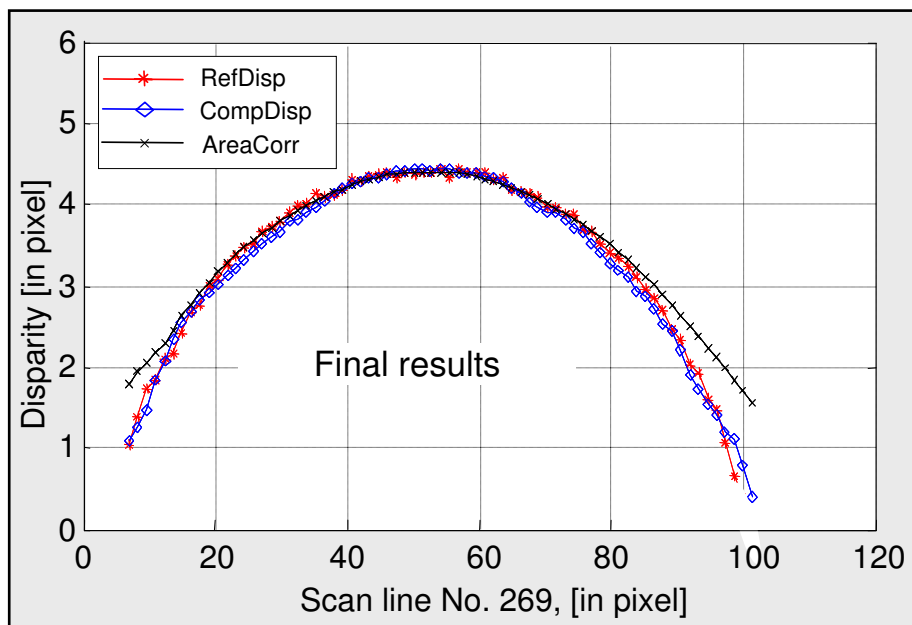


Figure 3.23: Comparison of the correspondence lines profiles from the computed disparity (CompDisp), the disparity from the Area-based method (AreaCorr) and the reference disparity (RefDisp).

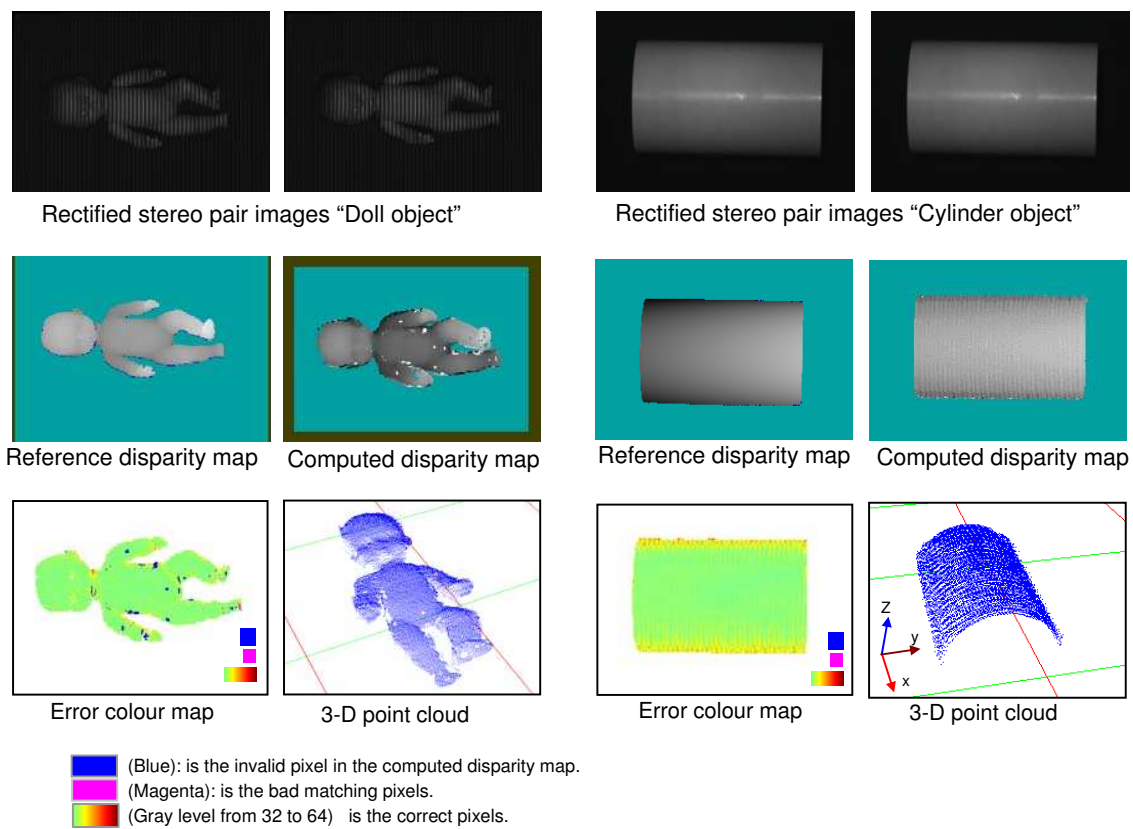


Figure 3.24: Two Examples of the final computed results of "Doll object" in left and "Cylinder object" in right. The stereo pairs (top), the reference disparity map and the refined computed disparity map (middle). The color error map and the 3-D point cloud for surface reconstruction (bottom).

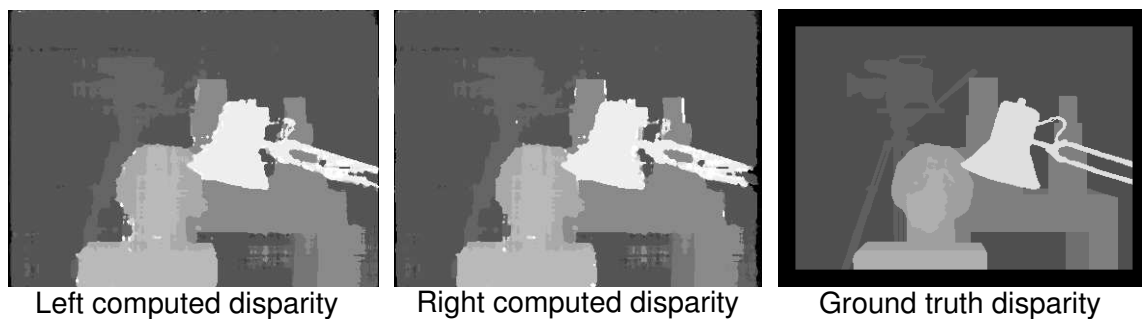


Figure 3.25: Refined disparity estimate to the same stereo pair shown in Fig. 3.18. Left and right computed disparity and its ground truth disparity map.

Chapter 4

Local Spatial Frequency Representation

4.1 Local Spatial Frequency

The space and frequency are the proper analysis domains in the stereovision algorithms. Therefore we would like to have an operator that analyzes signals simultaneously in both domains and provides information of localized space-frequency events.

On the use of a joint spatial-frequency representation, the 2-D imagery can be studied in a two views; a spatial view and the frequency view [61], [62]. In the spatial view, each individual pixel is paramount; an image is represented by the sequence of independent pixel values and each pixel is considered unique and important. In frequency view, the image is broken down mathematically into several frequency components, information in each component relates to the image as a whole. The combination between them is called *local spatial frequency*.

In phase-based methods, the stereo images are transformed to the frequency domain. Afterwards, the disparity is resulted as a phase difference between the image signals (image scanline). Many image phenomena are more succinctly described and more easily manipulated in the frequency domain than in the spatial domain. But several problems in traditional stereo arise from its limited image representation. Also the problem with using Fourier transform directly is that it extracts frequency information contained *everywhere* in the image, but you do not know where in the signal those frequency occur.

Two common approaches to image interpretation are frequency analysis (e.g., using Fourier transforms) and direct pixel-basis analysis (spatially dividing the image into a grid of pixels). Early only one of these approaches was used at a time; frequency analysis for global effect, or spatial analysis for local effects. But recently *local spatial frequency representations* allow both techniques to be used simultaneously, at the cost of greatly increasing the amount of data used to represent the image. Some examples of locale spatial representations are Wavelets, Spectrograms (Short Time

Fourier Transforms (STFT)), Wigner distribution (WD), and Scalograms (which are comprised of Gabor filters) [63]. All are similar in effect, but slightly different in structure. Space-frequency image representations characterize images over a space-frequency plane. They thus combine space-domain and frequency domain analysis to yield a potentially more revealing image of the temporal localization of a signal's spectral components.

Previous work with windowed Fourier transforms in computer vision reveals some of the potential utility of local spatial frequency analysis. Image spectrograms have been used for a variety of image analysis works.

All of these approaches use the Fourier transform over either the whole image or a fairly large region. The Fourier transform, however, hides the spatial coherence of the image. Thus, although it can identify the component frequencies of an image, their location in the image is ambiguous (this means that it is not possible to identify the local spatial structure of the image). Large-support Fourier transforms tend to smear the frequency peaks of signals whose frequency is changing (e.g. a periodic pattern on a tilted plane) and confound the analysis of signals with spatially distinct subcomponents (e.g. two adjacent textures). A solution to this problem is the space-frequency representation which shows the frequency content of only small, local regions of the signal.

Fig. 4.1 shows the idealized space-frequency representation. A sinusoidal wave in which the low-frequency wave with wavelength $\lambda = 20$ pixel (frequency = 0.05) occupies the center of the signal having $\lambda = 7$ pixel (frequency = 0.14) is shown in Fig. 4.1 (a). This signal is sampled into 128 pixels from each sinusoid fragment (we will call them image samples) for a total of 384 data point . The Fourier transform of this signal is shown in Fig. 4.1 (c). As a result of this transform, a pair of peaks appears, one peak for each frequency, but we can not determine *where* the sub-signals (the samples) occur in the space in the original signal. This is unacceptable for image matching. A 2-dimensional spectrogram representation is shown in Fig. 4.1 (d); the spectrogram is a two dimensional function of space (horizontal axis) and frequency (vertical axis). Which was computed directly from the image samples without preceding knowledge of the original continuous signals. The spectrogram is based on the short-time-Fourier-transform (*STFT*) with a fixed window size at all image samples. The window function's shape and width are chosen based on the task manually. Therefore, this representation has some practical and fundamental limitations, for instance, the Nyquist sampling theorem tells us that the analysis window must be at least twice as wide as the longest wavelength being analyzed, but a wide window will tend to blur the measurements around discontinuities in the original signal (observe the horizontal blurring around the dark lines in the spectrogram in Fig. 4.1 (d)). The structure of the signal is made clear in the local space-frequency scalogram representation (scalogram-phase) which is shown in Fig. 4.1 (b), scalogram illustrates a relatively higher-frequency component exists at the ends of the signal,

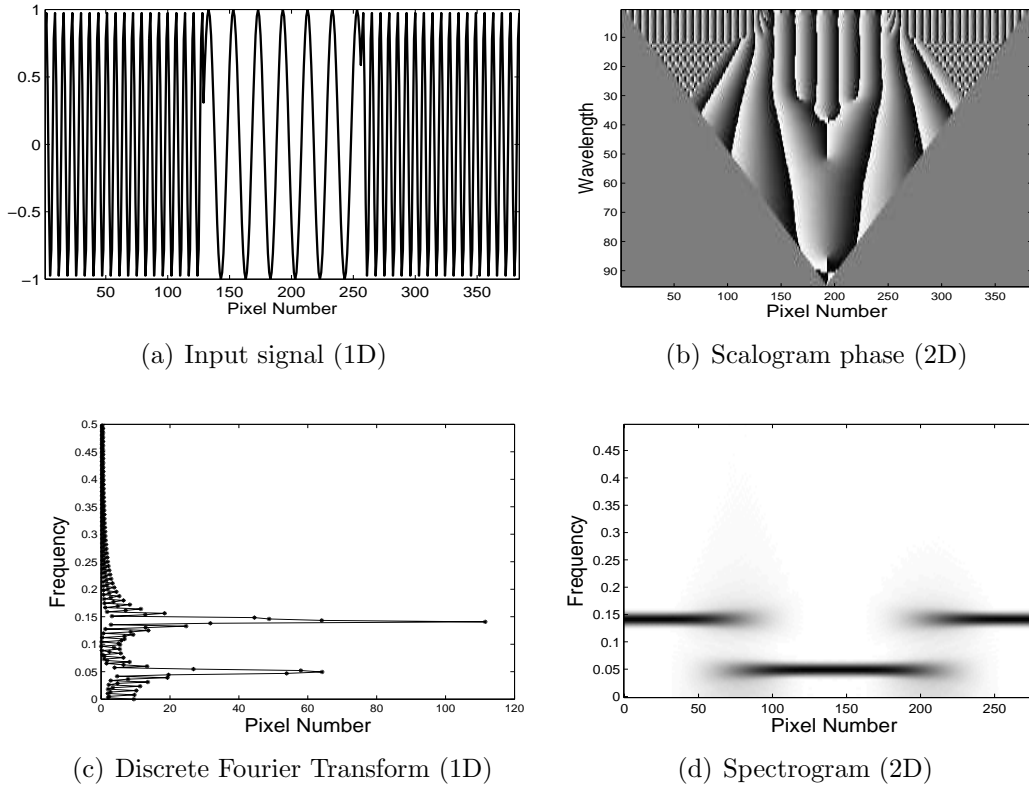


Figure 4.1: Local Spatial Frequency representation. (a) The one dimensional synthetic signal, a low frequency sin wave embedded in a high frequency wave. (b) The Local Spatial Frequency plot (Scalogram phase) associates each sample point with its proper frequency. (c) The Discrete Fourier Transform, which has two peaks, one for each frequency. (d) The Spectrogram representation, using a fixed kernel window without explicit knowledge of the original input signal's analytic form.

while a low-frequency part occurs in the middle. This localization is the power of scalogram representation, since the sampling is linear in wavelength.

4.1.1 Spectrogram

The spectrogram of a signal is a series of small-support. Fourier transforms of the signal, each centered around a different point of the signal. For a one-dimensional signal $f(x)$, the spectrogram is $SP(x, \omega)$ where ω is frequency in cycles/unit distance. The spectrogram is an estimate of the power of frequency ω at the point x . The continuous spectrogram of the one-dimensional function $f(x)$ is given by [64], [65];

$$SP(x, \omega) = \left| \int_{-\infty}^{\infty} W_l(a - x) f(a) e^{-j2\pi\omega a} da \right|^2 \quad (4.1)$$

where $W_l(x)$ is a window function with support length l . To calculate one vertical slice of the spectrogram for a given value of x , say x_0 , the signal is first multiplied

by a window offset by x_0 . This product is Fourier transformed where the magnitude is calculated from the complex values of the Fourier transform; and the non-negative half of the magnitudes serve as $SP(x_0, \omega)$, which is one column of the spectrogram. This process is repeated for every x .

The discrete version is computed using the discrete Fourier transform (*DFT*), which is discrete in both space and frequency. The window function controls how much of the rest of the signal contributes to the spectrogram at the point x . In terms of $W_l(\omega)$ and $F(\omega)$, the spectrogram can be written as;

$$SP(x, \omega) = \left| (e^{-j2\pi\omega x} W_l(\omega) * F(\omega)) \right|^2 \quad (4.2)$$

where $*$ is convolution [65]. The spectrogram uses a fixed window size at all scale and a logarithmic sampling of wavelengths. this can be useful for texture analysis, but seems less useful for image matching.

4.1.2 Wigner Distribution

One popular space-frequency representation is the Wigner Distribution (WD) [66] [64], for the use of quantum mechanics. Like the spectrogram, the WD produces a function of both space and frequency from a function of space alone. Practically speaking, the WD can effectively deal with signals whose frequency is changing, giving a clear indication of their instantaneous frequency.

For a one-dimensional function $f(x)$, the Wigner distribution is

$$WD(x, \omega) = \int_{-\infty}^{\infty} f\left(x + \frac{a}{2}\right) f^\dagger\left(x - \frac{a}{2}\right) e^{-j2\pi\omega a} da \quad (4.3)$$

In few words, the way to compute $WD(x, \omega)$ is to first calculate the product $f(x + \frac{a}{2})f^\dagger(x - \frac{a}{2})$, which is the original signal multiplied by a conjugated version of the original signal flipped around the point x . The \dagger denote as the conjugate and f^\dagger is the conjugated version of the original signal. The Wigner-distribution has been applied to texture segmentation and to shape from texture, but it is not good choice for image matching, that is because the cross terms that are introduced by the WD would make it complicate automated analysis, since it is even more difficult to distinguish the true frequency peaks.

Now, it is useful to ask what we want to see in the 2-D local spatial frequency plot. Of course we need not only the particular frequencies be known, but also the area in the original signal in which they occur. In contrast to the spectrogram, the scalogram uses a variable window size that proportional to the frequency at each pixel and provides to us what we need. The scalogram is a wavelet with a Gabor function as the transfer function.

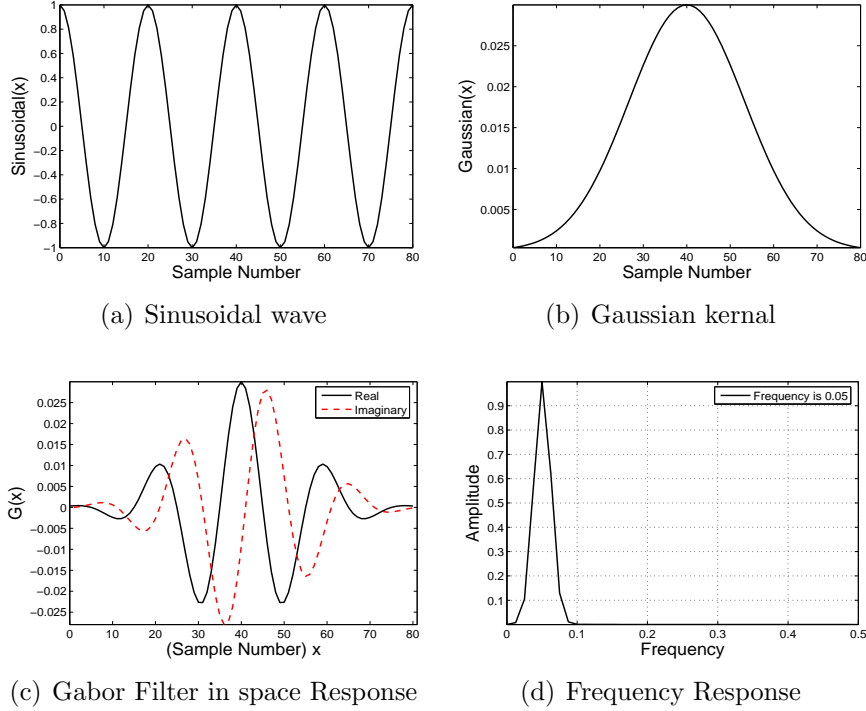


Figure 4.2: Gabor filter composition with tuning frequencies $\omega = 0.05$; (a) sinusoidal wave of particular frequency, (b) a Gaussian kernel, (c) the corresponding Gabor filter in a space domain, (d) Gabor filter in a frequency domain.

4.2 Gabor Filter

Gabor filters are the most commonly used tool in joint space-frequency uncertainty product and for the separability of center frequency and bandwidth [67]. The Gabor filter is a complex sinusoidal wave of particular frequency modulated by a Gaussian envelope which defines the space duration. As shown in Fig. 4.2. Gabor filters allow easy separation of the modulating component (which determines the spatial frequency) and the envelope (which determines the bandwidth).

Let $G(x, \omega)$ be the function defining a Gabor filter centered at the origin. For a given value of m and σ_f , we can view Gabor filter as:

$$G(x, \omega) = \rho(x, \omega).e^{-i\phi(x, \omega)} \quad \text{for } x \in \left[-\frac{m}{2\omega}, \frac{m}{2\omega}\right] \quad (4.4)$$

$$\text{with Magnitude; } \rho(x, \omega) = u.e^{-0.5\left(\frac{\omega x}{m\sigma_f}\right)^2} \quad (4.5)$$

$$\text{and Phase; } \phi(x, \omega) = 2\pi\omega x \quad (4.6)$$

where $u = \omega/(\sqrt{2\pi}m\sigma_f)$ is a scaling term, ω is the tuning frequency of the filter, m is the number of wavelengths to fit in the window, σ_f is a fraction of window size that

corresponds to one standard deviation of the Gaussian envelop σ , where $\sigma = \frac{m\sigma_f}{\omega}$.

An example of the Gabor filter can be seen in Fig. 4.2, the filter is presented with parameters $\omega = 1/20 = 0.05$, $m = 4$, and $\sigma_f = 1/6$. The frequency of the filter ($\omega = 0.05$) is clearly observed in the Gabor filter represented in frequency domain (see Fig. 4.2(d)). Gabor filters are closely related to the Fourier transform [34]. In fact, the Fourier transform uses the entire image to compute its results, the Gaussian window in the Gabor filter limits attention to a small region in the input image. In addition, unlike the Fourier transform which use a fixed window size at each frequency, the Gabor filter uses a window size that shrinks and grows as the tuning frequency changes (that is depending on the ω term in the Gaussian component of Eq. 4.4).

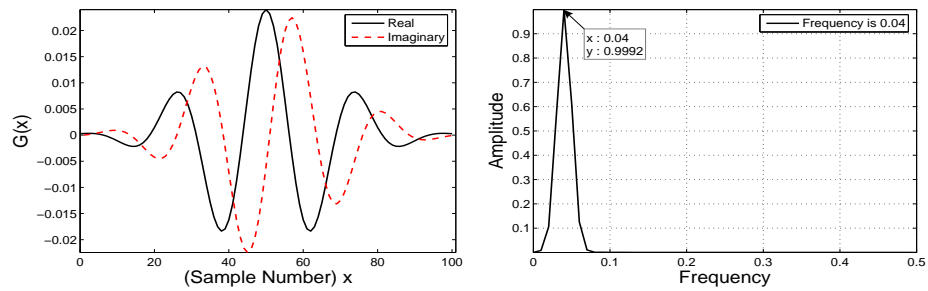
Fig. 4.3 illustrated the shrinks and grows of the window size as the tuning frequency changes. All of the filter in the figure have constant values of both the number of wavelength ($m = 4$) and the fraction of the window size ($\sigma_f = 1/6$). As we can observe, the effect space duration is inversely proportional to the effective bandwidth via the uncertainty relation. Moreover, the reason behind using the Gabor filter is, in Gabor filters, impulse response has the same support at low and high frequency. At high frequencies the effective window width will be quite small, ensuring that only the nearest pixels will be used to compute filter outputs. At low frequencies the window will be much wider, and will therefore require more data from the original signal.

Common to all Gabor features is that, they are based on Gabor filter responses for a given input image. The responses over the image are calculated for a set of filters, tuned to its parameters. Therefore, not only the tuning frequency is the important parameter in the Gabor filter but also the two other unexplained parameters (m and σ_f), which must be specified as the following:

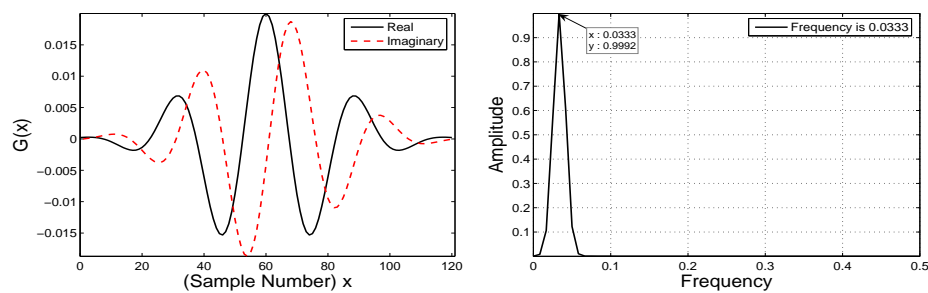
The first parameter is the number of samples, which is called, the number of wavelengths (m) to include in the window. The window size is equal to $m\lambda$ where λ is the filter wavelength, therefore, we are interested in looking at the question of how many samples should be taken so that no information is lost in the sampling process?.

The Nyquist-Sampling theorem establishes that "when sampling a signal, the frequency must be greater than twice the bandwidth of the input signal in order to able to reconstruct the original perfectly from the sampled version". So that, for an input signal whose lowest frequency component is ω_l (with corresponding to wavelength $\lambda_l = 1/\omega_l$), we must include $2\lambda_l$ samples of the signal to reliably extract information about that frequency. Consequently, we can choose that the value of m as long as $m \geq 2$. The default parameter value of the wavelengths per window that is used in this thesis is $m = 4$.

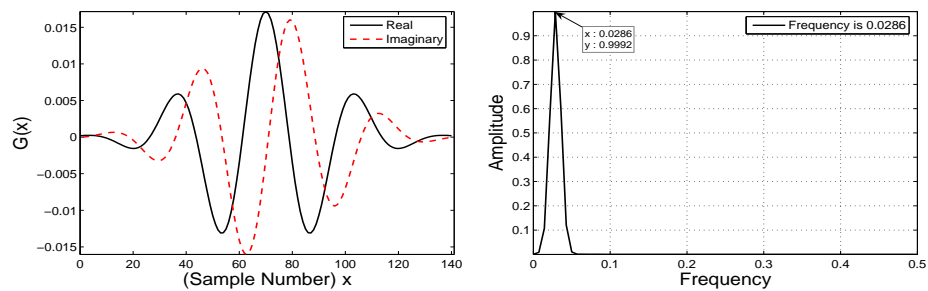
The second parameter is sigma fraction (σ_f). The parameter σ_f is the fraction of the window size that corresponds to one standard deviation σ of the Gaussian. For more explanation, we demonstrate how much of the Gaussian to include in the window. Theoretically, the Gaussian has infinite extend, but what is the reasonable



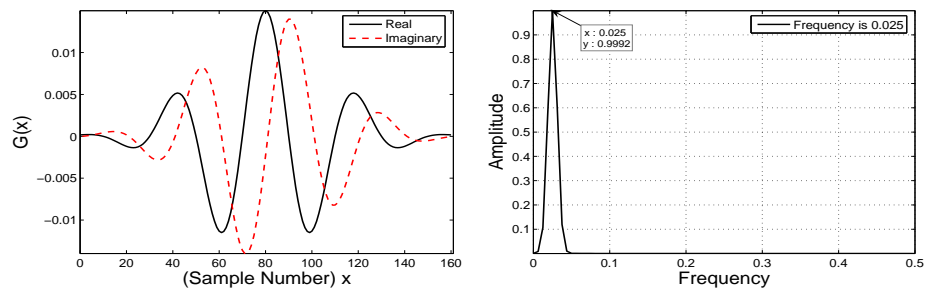
(a) Gabor filter in the space and frequency domains at tuning frequency $\omega = 0.04$



(b) Gabor filter in the space and frequency domains at tuning frequency $\omega = 0.0333$



(c) Gabor filter in the space and frequency domains at tuning frequency $\omega = 0.0286$



(d) Gabor filter in the space and frequency domains at tuning frequency $\omega = 0.025$

Figure 4.3: Gabor filter examples with constant values ($m = 4$ and $\sigma_f = 1/6$) but different values of the tuning frequencies (from top to bottom the ω equals 0.04, 0.0333, 0.0286 and 0.025). First column shows the Gabor filter in the space representation, and the second column shows their plots of frequency response.

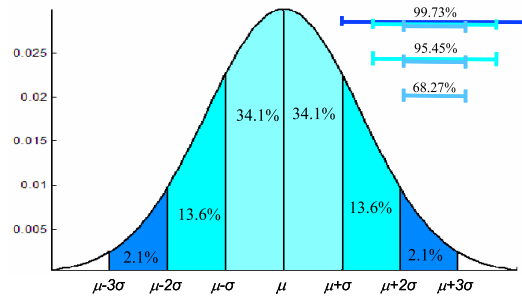


Figure 4.4: Gaussian distribution with mean μ , and three standard deviations of the mean (three on each side).

extent of the Gaussian window. In order to answer these query let us first look for the "Empirical rule" definition.

In practice, for a normally distributed data set, the *Empirical rule* states that 68.27% of the data elements (observation) are within one standard deviation of the mean, 95.45% are within two standard deviations, and about 99.73% lie within three standard deviations. Graphically, this corresponds to the area under the curve as shown in Fig. 4.4. Therefore, from this principle we can see approximately 99.73% of the observations fall within three standard deviations of the mean. The area under the curve over this range is the relative frequency of observations in the rang. Thus it is reasonable in practise to limit the extent of the Gaussian window to six standard deviations (three on each side of the mean).

In view of the fact that, Gabor is complex-valued, Fig. 4.5 illustrates two plots for a complete view of their components. The figure present Gabor filter as real-Imaginary pair in Fig. 4.5 (a), and as the Magnitude-Phase representation in Fig. 4.5 (b). The filter has a constant tuning frequency $\omega = 1/10$, and the values of the extra parameters m and σ_f are 4 and $1/6$ respectively. In Chapter 5, we will demonstrate the effect of the Gabor parameters on stereo disparity accuracy.

4.3 Gabor Scale-space Expansion: Scalogram

The scalogram is a local spatial frequency representations that make use of the local frequency content of an image. The scalogram uses a variable window size that shrinks and grows as the tuning frequency changes. It is actually a special case of the wavelet with a Gabor function.

For generation of the scalogram, $S_R(x, \omega)$ an adaptive Gabor filter convolve with a one dimension input row $R(x)$, caching the outputs in a two dimensional matrix with complex-valued elements (magnitude ρ and phase ϕ), that compute as follows:

Let $G(x, \omega)$ be the function defining a Gabor filter centered at the origin, as it

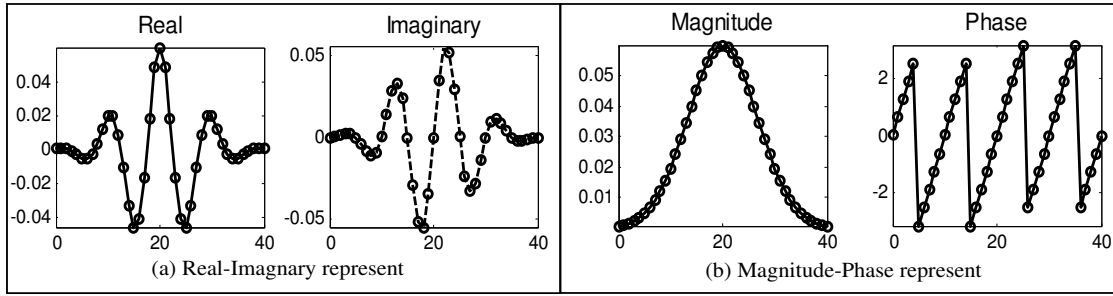


Figure 4.5: Gabor filter example with the tuning frequencies $\omega = 1/10$, number of wavelength $m=4$ and $\sigma_f = 1/6$. The Real-Imaginary pair, and the Magnitude-Phase pair are presented from left to right.

presented in Eq. 4.4 and $R(x)$ is the current image row.

Therefore, the scalogram representation i.e. the spatial convolution between $G(x, \omega)$ and the signal $R(x)$ is calculated as;

$$\begin{aligned} S_R(x, \omega) &= R(x) * G(x, \omega) \\ &= \int_{\tau=-\infty}^{\infty} R(\tau)G(x - \tau)d\tau \end{aligned} \quad (4.7)$$

Here the symbol $*$ denotes the convolution operation and the Gaussian envelope of $G(x, \omega)$ define the local neighborhood.

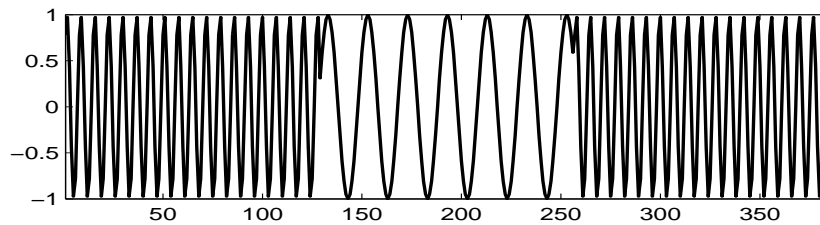
For our purposes, the convolution integral form, Eq. 4.7 (i.e. for the continuous signal) has been reformed as a sum of sample response (i.e. for a digital signal), so that, convolving the two vectors $G(x, \omega)$ and $R(x)$ is the same operation as multiplying the polynomials whose coefficients are elements of G and R . If a is the length of vector G and b is the length of vector R , then S_R is a vector of length $(a + b - 1)$ and whose x_0^{th} element is

$$S_R(x_0) = \sum_{j=1}^{\min(x_0, a)} G(j)R(x_0 + 1 - j)$$

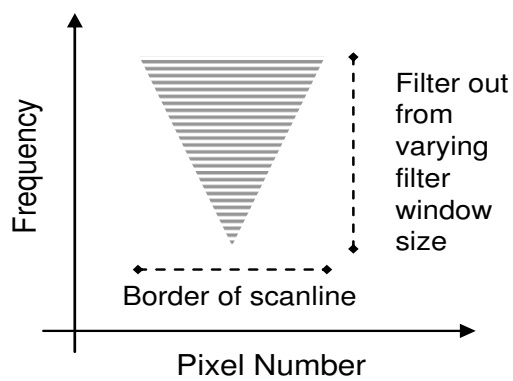
the sum is over all the values of j which lead to legal subscripts for $G(j)$ and $R(x_0 + 1 - j)$.

The sampling along the (vertical) frequency axis is one of the principal differences between the image scalogram representation and other local spatial frequency representations. This produces a multi-scale phase-based method, which can handle missing information at any scales. This kind of local spatial representation is very useful for *image matching* and *dense depth map reconstruction*.

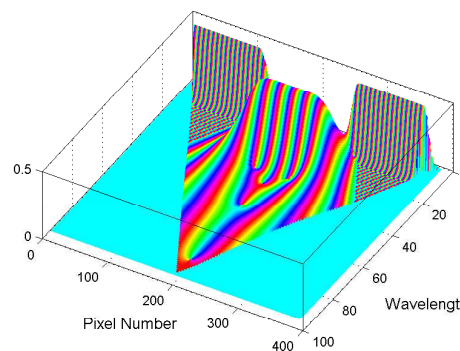
An example of the scalogram output for the input image scanline is shown in Fig. 4.6. A sinusoidal signal with different frequencies is shown in plot (a) as an



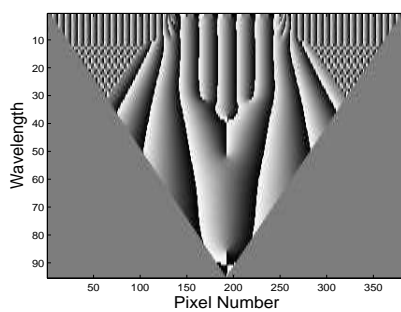
(a) Input signal



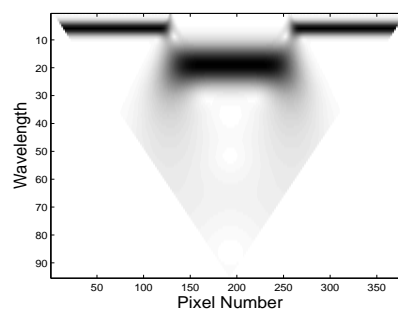
(b) Triangular shape for a scalogram representation



(c) Combined Phase and Magnitude



(d) Scalogram Phase



(e) Scalogram Magnitude

Figure 4.6: Magnitude and Phase Scalogram of a one dimensional sinusoidal input signal. The horizontal axis of the Scalogram correspond directly with the signal's horizontal axis. The vertical axes of the Scalogram correspond to a frequency scale with low frequency at the bottom.

input scanline. Since, the image scalogram is the result of applying a collection of Gabor filters to a 1-D image scanline and caching the output in a 2-D matrix, each representation is same like a triangular shape. The reason for this is due to the interaction between the varying filter window size and the border of the scanline. See Fig. 4.6 (b), since the filter window size proportional to the frequencies, the high frequency filter outputs found at the top of the scalogram are computed using only a few pixels, so their values can be computed over most of the width of the image. At lower filter frequencies, the number of points that must be sampled increases, reducing the number of outputs that can be computed. That means the sampling a long the vertical axis is linear in wavelength, which is reciprocal of frequency.

Fig. 4.6 (c) shows the combined scalogram-phase and scalogram-magnitude, the horizontal axis is the same as in the original signal (pixel numbers) and the vertical axis is wavelength (in pixels). The height of the plot encodes the strength of the signal at a given location and resolution (magnitude). The 2-D separated scalogram-phase and scalogram-magnitude for the input signal are shown in the bottom Fig. 4.6 (d) and Fig. 4.6(e).

Fig. 4.7 illustrates a synthetic input signal (R) and its scalogram representation. The phase plot of the scalogram is a particular interest because it enables us to actually detect the localization of the changing frequencies. At locations in the signal where there are large step changes, we can see a vertical line of constant grey value in the phase diagram (Fig.4.7 bottom) indicating a constant phase angle over all frequencies at that point in the signal. The arrows mark these vertical lines at the step transitions in the signal.

4.3.1 Phase-frequency Measurement

In this section, we discuss the response of applying the Gabor filter which is the sine wave multiplied by Gaussian to discrete signals (image scanline). The challenges arise in how measure the frequency of the sinusoid signal. In addition, the output of a Gabor filter is a complex number and in phase-based method phase plays an important role in the disparity computation, but how accurately can it be measured.

When filtering a periodic signal there are two common techniques for determining the frequency of the signal:

- using the frequency of the band-pass filter with the highest magnitude response
- measuring the phase derivative of the filtered signal

Using the filter tuning frequency directly is complicated because any discrete filter will have a blurred response, see the horizontal blurring around the dark line in the spectrogram representation in Fig. 4.1 (d). In this way, the frequency determined

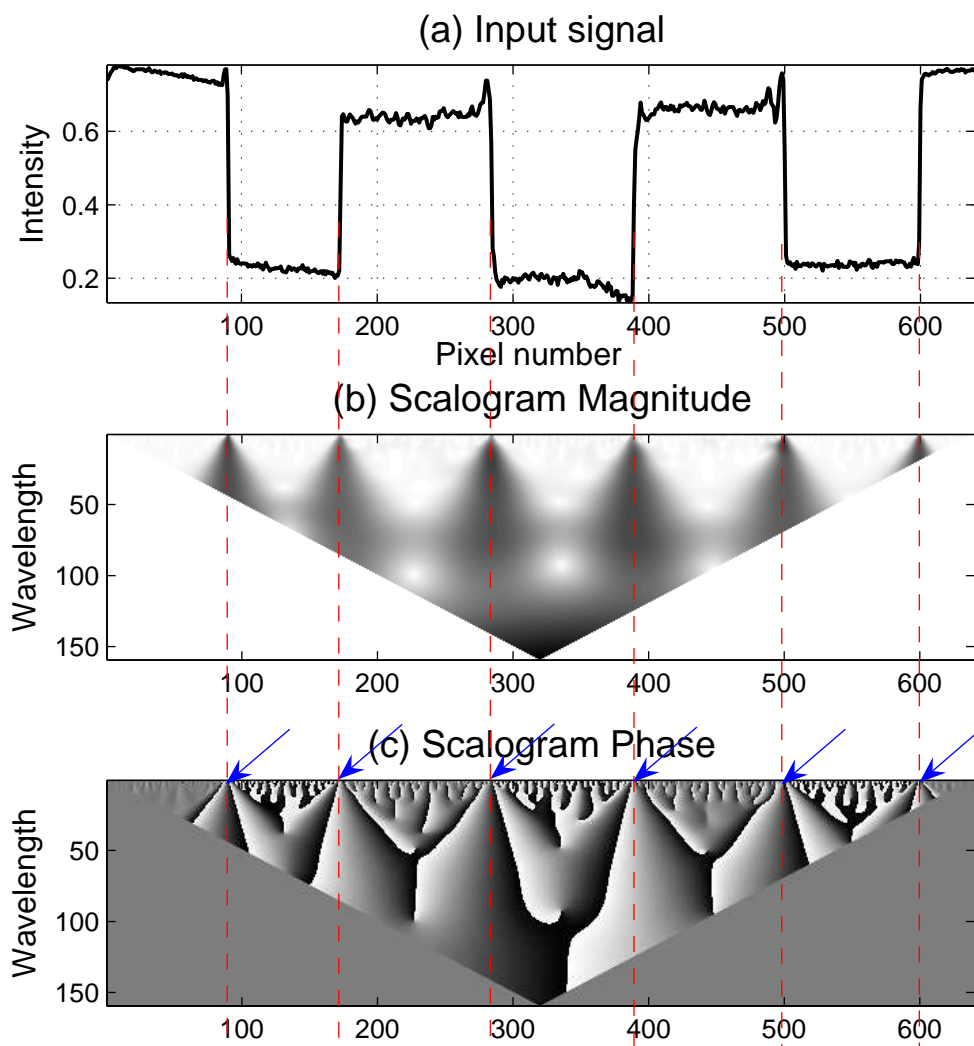


Figure 4.7: Discontinues view and its scalogram behavior: (a) Signal to be analyzed, (b) the magnitude scalogram, and (c) is its phase scalogram. In the phase plot the arrows mark vertical lines of constant phase that occur at step transitions in the signal.

by the convolution of the window with a bank of band pass filters tuned to different frequencies, find the filter with maximum magnitude response, and treat the tuning frequency of that filter as an estimate of the signal's instantaneous frequency.

The phase derivative that is known as the instantaneous frequency provides a more accurate measure of a sinusoid's frequency in theory, in practise its accuracy depends on the amplitude of the input signal. This has led several researchers to develop a constraint that filters out unreasonable phase-derived frequencies [3]. Similarly we provide in the next chapter a useful combined constraint which uses the phase and magnitude notation.

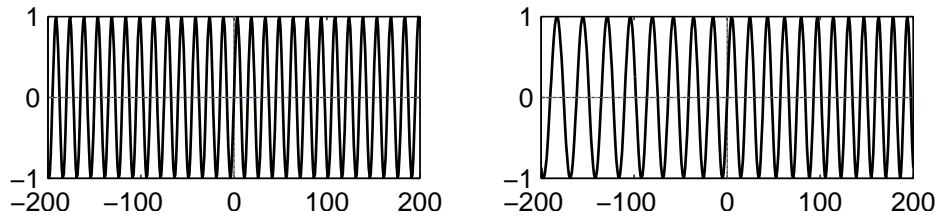
In order to understand how the instantaneous frequency can be computed (i.e. how the frequency and phase are related), we will demonstrate the same analysis of two different input sinusoids (as the input scanline) the first sinusoid with fixed frequency, $\omega = 1/15$, and the second one with varying frequencies from $\omega = 1/30$ to $\omega = 1/15$, shown in Fig. 4.8 plot (a), and a band pass filter (a single Gabor filter) with a tuning frequency $\omega = 1/20$. Plot (b) shows the derivative of the phase of the input signals, as we can observe in plot (b) the value of this derivative is everywhere equal to the frequency of the original signal. By looking at the original signal and considering their known frequencies to analysis what we see along the x-axis in the two plots (a) and (b) in Fig. 4.8. A sinusoid has a fixed shape, but may be stretched or compressed along the x-axis by changing it's frequency (e.g. the second sinusoid signal in this example). Similarly, the phase of a sinusoid will always change linearly, and its rate of change will be related to the original's frequency.

The linear change in phase of a sinusoid can also be measured in an arbitrary signal using a band pass filter, which can be an original sampled signal with a Gabor filter whose tuning frequency $\omega = 1/20$. Fig. 4.8(c) shows the phase which is computed by the Gabor filter (in a solid line) is approximately equal to the analytically known phase of the original signal (in a dashed line), the phase wrap around at 2π exactly where one period of the sinusoid repeats.

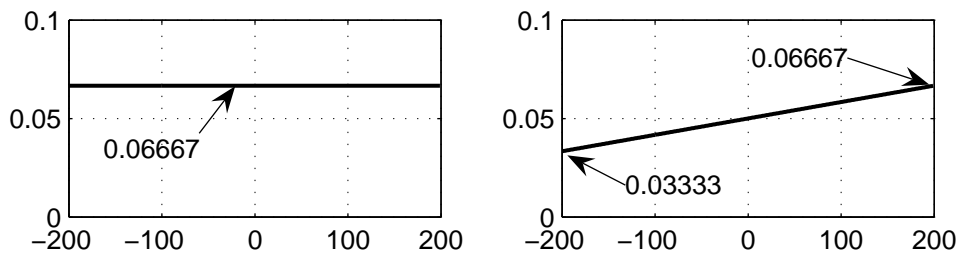
Finally, the derivative of the Gabor response phase is shown in Fig. 4.8 (d), again with the analytically known value of the original signal's frequency (phase derivative). We can observe that, except around the discontinuities at the ends of the signal, the empirically-derived Gabor estimates are extremely close to the analytically known values, in particular when the frequency keeps changing.

The most interesting property to notice here is that this method of computing instantaneous frequency is reasonably independent of the analysis filter's tuning frequency, since the same Gabor filter was used to compute the phase in the both different original signals (in the previous example). But the accuracy depends on the amplitude of the input signal, therefore section 5.2.2 in chapter 5 develops a constraint to filter out unreasonable phase-derived frequencies.

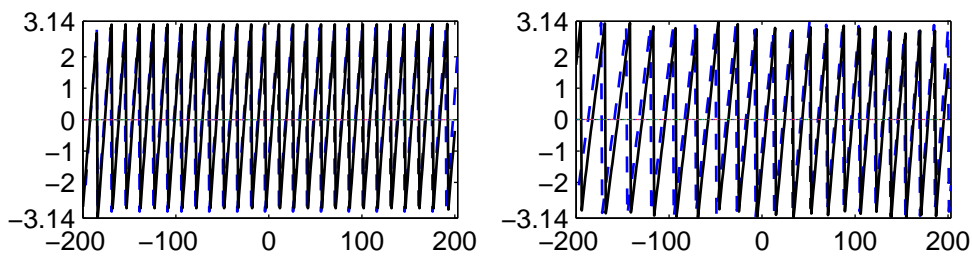
Generally, in the phase-based stereo matching method, in order to compute a precise disparity estimate, it must reflect measurements taken at the same frequency,



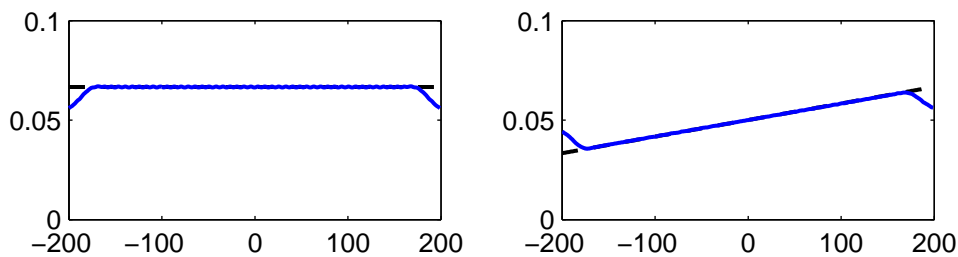
(a) Original signals, First signal ($\sin 2\pi\omega x$) with a fixed frequency $\omega = 0.06667$ and a second signal $\sin(2\pi x(\frac{1}{20} + \frac{x}{24000}))$ with frequency varying from $\omega = 0.03333$ to $\omega = 0.06667$



(b) Phase derivatives, $(\frac{1}{2\pi} \frac{d\phi(x)}{dx})|_{2\pi}$ for the input original signal



(c) Phase of Gabor responses, approximately equal to the phase original



(d) Phase derivative of Gabor responses (in solid line) against the known value of the phase derivative of the original signal

Figure 4.8: Instantaneous frequency analysis for two signals. The signal with fixed frequency (left) and signal with varying frequency (right).

but what if the instantaneous frequencies measured by same filter on a pair of corresponding image regions differ, as can occur with perspective foreshortening. One of our contribution provided in chapter 6, manipulate this situation by providing a foreshortening correction factor based on the physical geometry of the scene.

Chapter 5

Phase-based Disparity Estimation

In this chapter we restrict our attention to a particular class of phase-based method, the so-called phase-difference-based stereo method [7], [8]. In the previous chapter, we give an overview of local spatial frequency, develop our method, and contrast it with existing ones. The method that is developed here will be analyzed and extended in the next chapter to address issues of foreshortening problem.

The phase-difference-based technique has become a widespread method for depth and optical flow estimation because of its superior performance and better theoretical grounding. The technique is based on the convolution of the stereo image pair with Gabor filters. Gabor filter contains two parameters, the width and the tuning frequency. In order to optimize its performance, these parameters have to be chosen in accordance to the characteristics of the visual signal. In this chapter, we propose an automatic technique to locally adapt the filter parameters to the input signal. At the beginning of the experimental analysis, we analyze the performance of the phase-difference-based technique for disparity estimation with respect to the choice of the Gabor filter parameters. In particular, we characterize the effects of phase nonlinearities on the quality of disparity estimates. In the second part, a novel technique is introduced which reduces phase nonlinearity by means of an adaptive mechanism for the tuning frequency. The performance improvement that is produced by the adaptive filter is demonstrated using different types of images. Results show that the proposed technique allows a significant improvement of disparity estimation.

5.1 Phase-difference Based Method

5.1.1 Phase-difference as Disparity

As an alternative to spatial correlation, phase based methods of disparity measurement have been proposed by a number of researchers [2], [41], [30], [34]. The disparity is estimated from local phase-difference using the instantaneous frequency, where the instantaneous frequency is estimated from the derivative of the local phase.

Phases	ϕ_l	0	0	0
	ϕ_r	0	$-\pi/4$	$-\pi/2$
Disparity		0.0	2.5	5.0

Table 5.1: Application of Equation 5.2. Estimating disparity as phase difference

At the first, it is useful to start from the principle mathematical foundations with a simplest possible example. In stereovision, the simplest case is typically the comparison of two 1-D functions that represent scanlines from the rectified stereo images. In the following example (Fig. 5.1, we are interested in studying how the phase of a sine wave relates to stereo disparity.

Let the left and right image $F(x_l)$ and $F(x_r)$ be sinusoids wave, which are shown in Fig. 5.1 (right), each of them have a particular phase angle, θ ;

$$F(x_l) = \sin(2\pi\omega x_l - \phi_l), \quad F(x_r) = \sin(2\pi\omega x_r - \phi_r) \quad (5.1)$$

where ω is the a frequency, ϕ_l and ϕ_r denoted as the phases of the left and the right images. The disparity is the amount of shift required to make the left and the right images appear equal. Mathematically, we discover the disparity by setting the formulas in Eq.5.1 equals.

$$\sin(2\pi\omega x_l - \phi_l) = \sin(2\pi\omega x_r - \phi_r)$$

$$2\pi\omega x_l - \phi_l = 2\pi\omega x_r - \phi_r$$

Then the disparity at any point in the signal is a different distance at that position ($x_l - x_r$) given by;

$$Disparity = x_l - x_r = \frac{\phi_l - \phi_r}{2\pi\omega} \quad (5.2)$$

Graphically, Fig. 5.2 shows the disparity as a horizontal separation between the two signals. In this example, all pixel disparities will be considered equal, let the frequency $\omega = 1/20$, the phase of left image $\phi_l = 0$ and allow the right phase ϕ_r to vary, (e.g. $\phi_r = 0, -\pi/4, \text{ and } -\pi/2$). Then the disparity is measured as the amount of shift visible in the graphs. When the right phase equal to 0 the two signals are matching together and the disparity equal to 0, in case of the right phase equal to the $-\pi/4$, we can observe the amount of horizontal shift is 2.5 units. Similarly, when $\phi_l = 0$ and $\phi_r = -\frac{\pi}{2}$ in Fig. 5.2 (bottom) we can see the horizontal separation between the two signal is 5 units.

Furthermore, by plugging the value of ϕ_l and ϕ_r into Eq. 5.2, we can compute the disparity values, the obtained result which are found in Tab. 5.1 equal to the amount of shift, that are indicated by the labeled bar in the Fig. 5.2.

This example also give us an important observation of the "phase-wraparound problem" in the phase-difference based method. This problem is caused by the fact

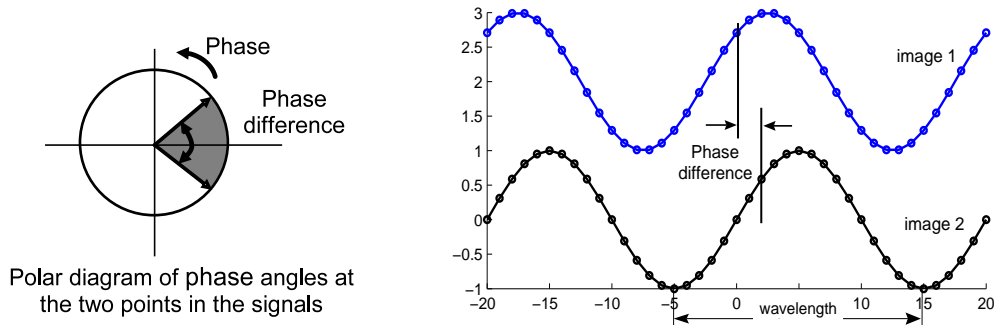


Figure 5.1: Estimating disparity between two images from phase difference and instantaneous frequency.

that the difference between 359° and 1° is 2° , not 358° as calculated by a simple subtraction. On the other wards, in the example of Fig. 5.2 where the frequency $\omega = 1/20$, disparities of 1, 21, 41, ... all appear equivalent. This means, we can not compute a unique disparity from a single phase value. Some authors [4], [3], [34] attempt to address this problem. More details and our suggested solution for the wrap-around problem will be demonstrated in section 5.2.3.

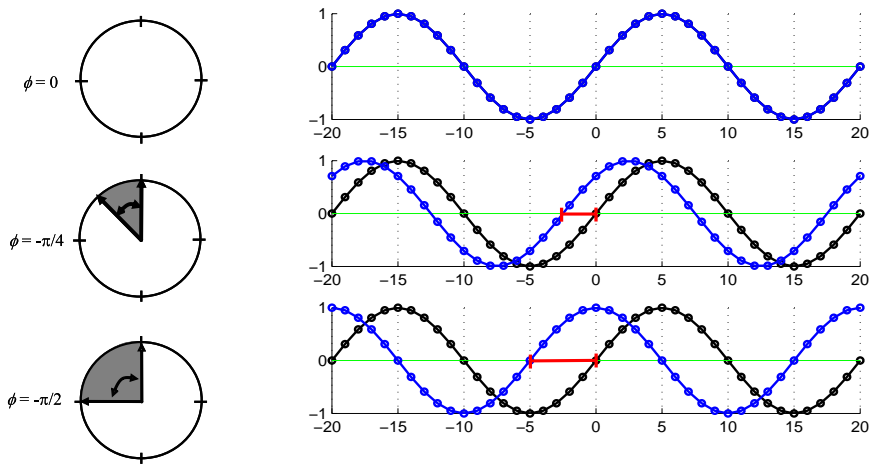


Figure 5.2: Disparity as a function of Phase difference. The disparity is a horizontal separation between the two images.

Now, it will be more convenient to expose the mathematical derivative of that basic idea of the disparity measurement as a phase-difference. The heart of phase-based methods is the filters that decompose the image into band-pass signals. Let us suppose that the arbitrary disparity Δx is constant over the image as shown in Fig.5.2, this is a global disparity that will apply to every pixel in the image. Then, according to the Fourier shift theorem, which states that a translation Δx in space corresponds to a phase shift $2\pi\omega\Delta x$ in the spatial frequency domain:

$$f(x - \Delta x) \xrightarrow{F} F(\omega).e^{-j2\pi\omega\Delta x}$$

where F is the Fourier transform of the image $f(x)$.

The fundamental idea of the phase-based algorithm to disparity measurement is to recover the disparity as a phase difference observed in the frequency domain. Let the $l(x)$ and $r(x)$ be the left and right images. They are related to each other by shift $r(x) = l(x - \Delta x)$. Then the Fourier transform for each one is computed independently

$$l(x) \xrightarrow{F} L(\omega)$$

$$r(x) \xrightarrow{F} R(\omega)$$

Moreover, it is assumed that the two signal/image are related by a shift Δx , and according to the Fourier shift theorem

$$r(x) = l(x - \Delta x) \xrightarrow{F} L(\omega).e^{-j2\pi\omega\Delta x}$$

Then the Fourier transforms of the image pair $L(\omega)$ and $R(\omega)$ are related to each other by

$$R(\omega) = L(\omega).e^{-j2\pi\omega\Delta x} \quad (5.3)$$

This equation holds for any value of frequency ω , therefore all of the complex Fourier coefficients from the right image are exactly the same as those for the left image, except that their phases are shifted by some amount.

The polar representation of the real and imaginary parts in the complex plane is denoted as, magnitude ρ and phase ϕ , where the magnitude $\rho_r = |R(\omega)|$ and the phase denotes the argument of the complex response $\phi_r = \arg [R(\omega)]$. Since the magnitude of the coefficients does not change, let us now restrict our attention to the phase components of Eq. 5.3. Using this notation, Eq. 5.3 written as:

$$\arg [R(\omega)] = \arg [L(\omega).e^{-j2\pi\omega\Delta x}]$$

$$\phi_r = -2\pi\omega\Delta x + \phi_l$$

Therefore, the disparity, Δx , is obtained as a function of phase difference;

$$\Delta x = \frac{\phi_l - \phi_r}{2\pi\omega} = \frac{\phi_l - \phi_r}{2\pi} \cdot \lambda \quad (5.4)$$

It is the same results that presented in Eq. 5.2, where Δx denoted as the disparity value.

Given a frequency ω_0 the disparity Δx is computed using the difference of phases $\phi_l(\omega_0)$ and $\phi_r(\omega_0)$. This is correct in case of having a fixed frequency over all the signal (the image scanline), e.g. the continuous sine wave example. But unfortunately, the recovered disparity is global and can not be assigned to a particular region in the image. Furthermore Eq. 5.4 is assumption that the left and right phases are measured at the same frequency and that is not reasonable in the real images.

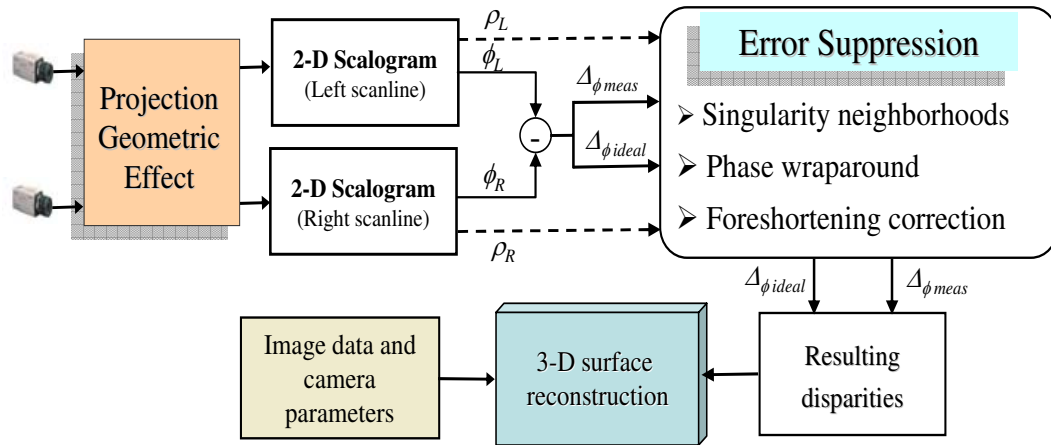


Figure 5.3: Concept of phase difference-base suggested algorithm. Besides on the phase and magnitude information of the rectified input stereo images using the measured-phase difference $\Delta_{\phi_{meas}}$ and the ideal-phase difference $\Delta_{\phi_{ideal}}$

5.2 Suggested Phase-difference Based Algorithm

Stereovision is a technique used to extract depth from a pair of images taken from slightly different view points. Depth can be computed from the relative position of corresponding points in the two images. Therefore, to have a local disparity estimation, it is necessary to define a window inside which the points are picked up to compute the phase terms. Hence, we propose to use a Gabor scale space expansion (scalogram) with adaptive window size to capture both local orientation and frequency information from the image.

Fig. 5.3 shows the concept of our suggested phase-difference algorithm. It will be described in several levels; the primary step is the preprocess image rectification. In order to simplify the matching process, rectifying the images is one of the important and first step in our method. By using the camera parameters that are obtained from the camera calibration process the captured stereo images are transferred to images as if obtained by camera with parallel optical axis. Therefore corresponding points must always lie along epipolar lines in images. These lines correspond to the intersections of an epipolar plane (the plane through a point in scene and nodal points of the two cameras) with the left and right image plains. exploiting this epipolar constraint reduces an initially 2-D search to a 1-D one. Thereby the first level deals with the image representation in the local spatial representation and the component features extraction (phases and magnitudes). The second processing level of the approach is specified by exactness the complementary problems (singularity neighborhoods,

wraparound, and the precision effects of the foreshortening). Finally, the 3-D coordinates of the examined object points are obtained using conventional triangulation method. In the following subsections we will discuss these complimentary problems as well as the program and its implementation requirements.

5.2.1 Image Scalogram

Image matching is important for 3-D stereo surface reconstruction. In this task, matching is found by shifting one image to match the other; the amount of shifting needed at each point reveals the 3-D structure of the scene. If the portion of the image is uniform with no features, then matching is impossible. If features are present, a match can be obtained. Here, as in other spatial vision tasks, it will be more convenient to work with the image spatial-frequency representation (image scalogram representation). Image scalogram is the result of applying a collection of Gabor filter to a 1-D image scan-line. The matching precision available at any point in the image is limited by the highest frequency present at the point.

Fig. 5.4 illustrates the brightness of a 1-D image scan-line, X_l , and its corresponding scalogram representation (phase and magnitude). The scalogram plots have a straightforward interpretation; the horizontal axis is the same as the original signal (pixel number) and the vertical axis is linear in respect to wavelength (in pixels). In the scalogram plot we observe that the short wavelength are on top; while the longer wavelengths are at the bottom. The intensity of the points is the scalogram-magnitude (see scalogram-magnitude plot) that encodes the strength of the signal at a given location.

It is easy to see the relative contribution of each frequency to each point in the image's line-profile, the frequency information is very well localized. But in some regions in the scalogram phase, we can observe the phase signal has an unwanted behavior, as well as the same regions in the scalogram magnitude, which are weak and have value zero or nearly to zero. These points appears in white spots in the scalogram magnitude. At these points the disparity estimate is not computed accurately, this problem known as the *singularity points*.

5.2.2 Recognizing Singularity Neighborhood

The central and main advantage of phase-difference technique is the expected stability of phase through scales with respect to contrast differences between the left and right views, and its linearity through space. In some regions, the phase signal has an unwanted behavior. This is because, the Scalogram is analytic, and contains a number of isolated zeros, when $S_R(x, \omega) = 0$ and the magnitude is generally weak. As a result the disparity will not be computed accurately.

From the scalogram-phase and scalogram-magnitude plots in Fig. 5.4, we can observe that in some regions the phase scalogram has an unwanted behavior (i.e., the

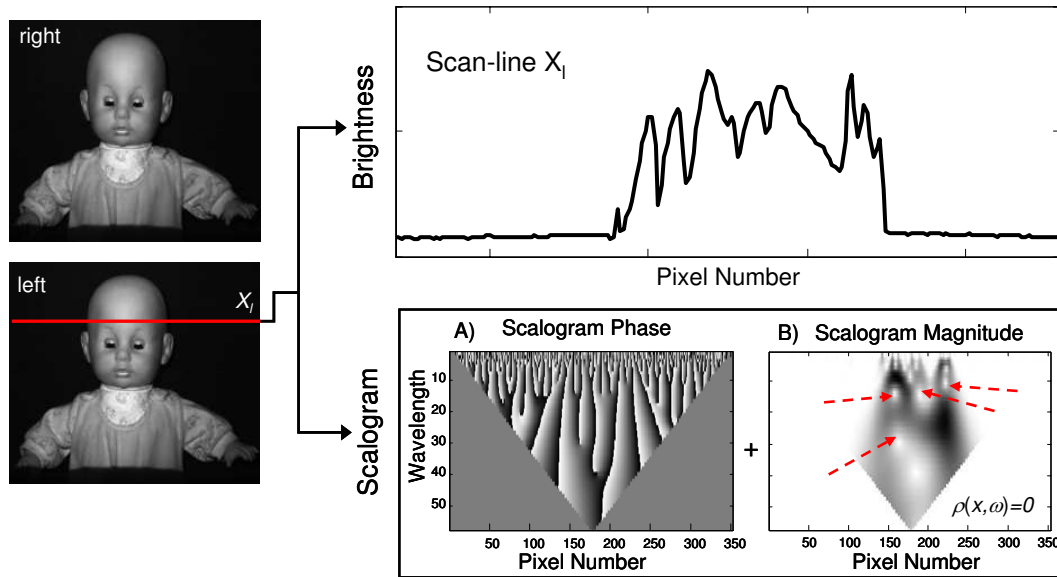


Figure 5.4: Shape of the signal from a complete scan-line and the scalogram representation of this signal (both magnitude and phase).

phase stability is not uniform throughout scale-space) and phase contours are nearly horizontal and not vertical as desired.

Moreover, by considering the interesting regions in the scalogram representations, Fig. 5.5 shows zeros appears as white spots that are marked by the arrows in scalogram-magnitude (Fig. 5.5 top-left). The phase signal is undefined when the magnitude fades away (these points known as the *phase singularities*). The expected density of these singularities are proportional to the peak tuning frequency [3], [30]. Fig. 5.5 (mettle column) illustrates the relationship between low magnitude and non-linear phase. The top figure shows level magnitude and its correspondence phase contour shown in the bottom. The scale for $\lambda = 7$ is marked by dashed line over the magnitude and phase contours are plotted in the left column of the figure, that show the local frequency (peak tuning frequency) and the phase as a function of spatial position (Note that, the marked dashed line crosses two singularity neighborhoods).

Now, we describe the characteristic behavior of $\rho(x, \lambda)$ and $\phi(x, \lambda)$ near singular points. Let $P(x_0, \lambda_0)$ denote the location of a singularity where the singularity point lies near the center at the point which the phase contours intersect, see Fig. 5.6 (B). The neighborhoods just above singular points (i.e., for $\lambda > \lambda_0$) are characterized by local frequencies that are significantly below the corresponding peak frequency ($\frac{2\pi}{\lambda}$). Within these neighborhoods regions exist in which local frequency is zero; i.e. $\frac{\partial \phi(x, \lambda)}{\partial x} = 0$. At this region, the level phase contours are horizontal and not vertical as expected, consequently the phase matching will be very sensitive to small changes in scale. Likewise, lower the singular point (i.e., for $\lambda < \lambda_0$) the neighborhoods are

characterized by local frequencies, which are significantly higher than corresponding peak tuning frequencies. Furthermore, the neighborhoods to left and right of singularity can be characterized in terms of magnitude variation which as $\rho(x, \lambda_0)$ goes to zero.

Singularity Neighborhood Detection

In order to increase the accuracy of disparity estimate, it is necessary to detect the singularity neighborhoods and avoid the use of these regions from the calculation. Fleet et al. [3], [30] have proposed two constraints to remove such areas. As similarly, we describe constraints that used to identify singularity neighborhoods without requiring the explicit localization of the singular points. The first constraint in Eq. 5.5 detects the neighborhoods above and below the singular points. The constraint removes all the areas where the local (instantaneous) frequency is deviated from the tuning frequency. In particular, we constrain the distance between local frequency of response $\frac{\partial\phi(x,\lambda)}{\partial x}$ and the peak frequency $\frac{2\pi}{\lambda}$

$$\frac{1}{\sigma_f(\lambda)} \cdot \left| \frac{\partial\phi(x, \lambda)}{\partial x} - \frac{2\pi}{\lambda} \right| < \tau_\phi \quad (5.5)$$

where $\sigma_f(\lambda) = \frac{1}{\sigma(\lambda)}$ and $\sigma(\lambda)$ defines the radius of the filter support.

The neighborhoods to the left and right of the singular points can be detected with the second constraint on local magnitude variation, Eq. 5.6, i.e. where the magnitude is very weak.

$$\sigma(\lambda) \left| \frac{\partial\rho(x, \lambda)}{\partial x} \right| \cdot \frac{1}{\rho(x, \lambda)} < \tau_\rho \quad (5.6)$$

$\rho(x, \lambda)/\frac{\partial\rho(x,\lambda)}{\partial x}$ approximates the distance from the singularity to x with a linear magnitude model [3]. For the fact that, near the singularity the magnitude acts like a parabola (see Fig.5.6 (B)), we can observe that the level magnitude contours near the singular point drawn as a parabola, it follows that the constraint in Eq. 5.6 removes the boule of the area size.

Now, we suggest combined constrain which uses the phase and magnitude notation for a Gabor filter with tuning wavelength. Consequently, for finding reasonable measurements we can express a combined stability constraint as;

$$\sigma(\lambda) \left| \frac{\partial\rho(x, \lambda)}{\partial x} \cdot \frac{1}{\rho(x, \lambda)} + \frac{\partial\phi(x, \lambda)}{\partial x} - \frac{2\pi}{\lambda} \right| < \tau \quad (5.7)$$

As τ decreases, this constraint detects larger neighborhoods around the singular points. This formula was derived from two observations; the first is "the phase derivative of Gabor filter should be roughly" and the second is "the magnitude derivative should be small".

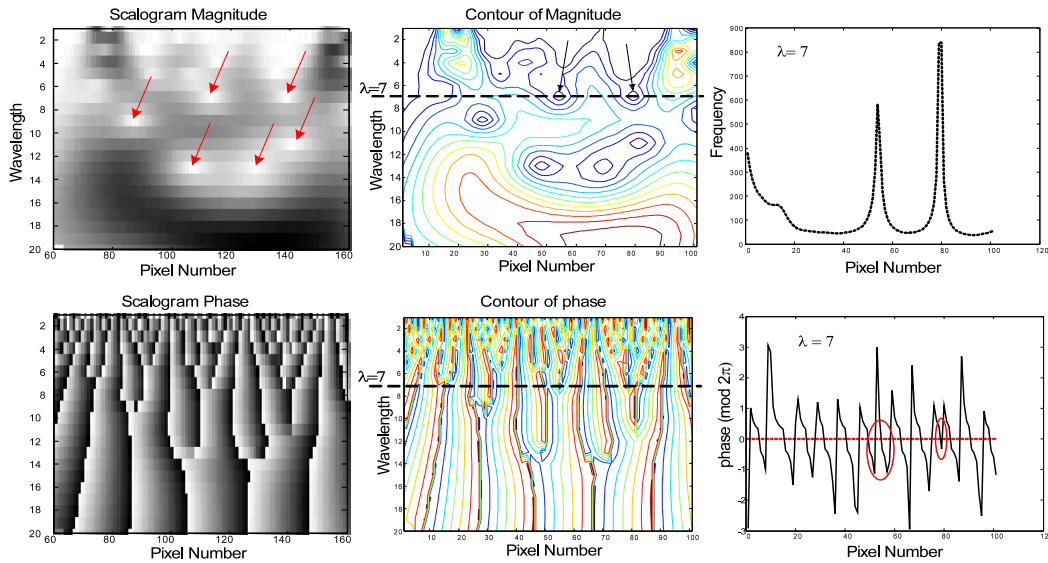


Figure 5.5: Detection of singularity neighborhoods. The scalogram response (magnitude and phase components) of a line-profile from figure 5.4, with $1 \leq \lambda \leq 20$. The two plots in left column are the magnitude $\rho(x, \mu)$ and phase $\phi(x, \lambda)$. In the middle column, the level contours of the magnitude on top and phase in bottom. Right column is the local frequency and the phases as a function of spatial position at $\lambda = 7$.

Fig. 5.7 shows the application of Eq. 5.7 at $\tau = 0.06$. The detected regions of the singularity neighborhoods that achieve the constraint comes into view as labeled regions in the level phase contours Fig. 5.7 (A), as well as in the level magnitude contours in Fig. 5.7 (B). Therefore, the unreliable phase values which labeled by these regions will be ignored.

5.2.3 Combined Algorithm for Estimating Disparity

Our algorithm computes the disparity at each pixel using phase-differences. In section 5.1.1, we demonstrate the principal of phase-difference technique for disparity estimates, and how the phase of a sinusoid could be used to compute stereo disparity. Therefore, our technique based on the convolution of the stereo image pair with several Gabor filters, and caching the outputs in 2-D image scalogram (magnitude and phase). In the beginning, we exploit the front-parallel surface assumption that assumes that the depth is constant over a region under consideration. This description will be extended in chapter 6 to consider the case in which slanted surface (foreshortening problem).

Let a pair of images I_L and I_R be given. Assume that these images have the same dimensions, and were taken with a pair of cameras with proper calibration, this

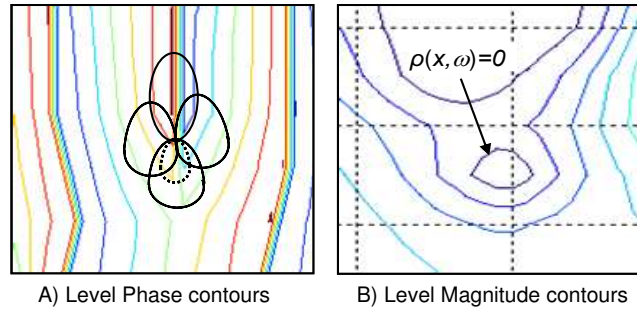


Figure 5.6: Characteristic behavior of level phase and magnitude contours in singularity neighborhoods. A) the small ellipsoidal contour marks location at which $\frac{\partial \phi(x, \lambda)}{\partial x} = 0$, the other ellipsoidal contours illustrate the regions detected by the constraints. B) the level magnitude contours near singularity that appear like a parabola.

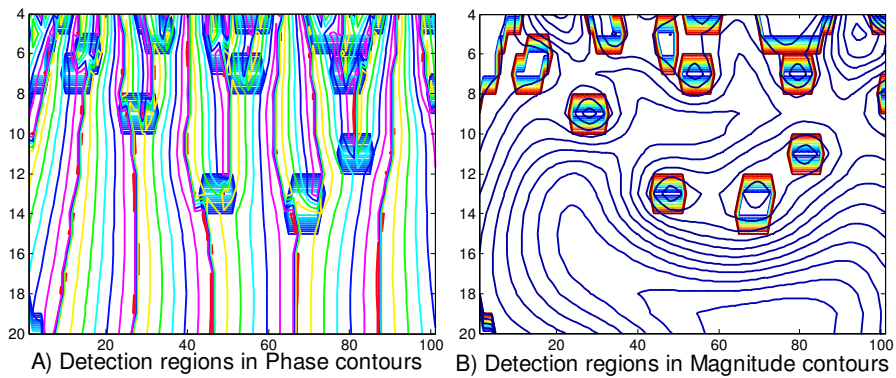


Figure 5.7: Application of stability constraint. Detection of singularity neighborhoods by the combine constrain 5.7 for $\tau = 0.06$ and $4 \leq \lambda \leq 20$. A) level phase contours and the region which detected shown in the labeled regions. B) The correspondence Magnitude contours that achieve the constraint.

constraint reduces the matching problem from two dimension to one, that is because same-numbered of scan-lines in the two images are guaranteed to correspond to the same plane in the world. Then for each corresponding rows x_l and x_r in the pair images, each contains n pixel intensities. At first, according to the characteristics of the scan-lines the scalogram is computed as a result of convolving a collection of Gabor filters with a 1-D image scan-line. The scalogram consists of a 2-D matrix of complex values represented by; magnitude ρ and phase ϕ . Where, $\rho_l(c, \lambda)$ is the magnitude of the left row's scalogram entry at pixel c with wavelength λ , and $\phi_l(c, \lambda)$ is the phase of the left row's scalogram entry at that pixel. Similarly for a right row with magnitude $\rho_r(c, \lambda)$ and phase $\phi_r(c, \lambda)$.

For a given pixel c , we look at the corresponding columns in the magnitude plots $\rho_l(c, \lambda)$, $\rho_r(c, \lambda)$ see Fig. 5.8(c). These columns represent the strength of the signal at many wavelengths, Fig. 5.8(c) shows the scalogram-magnitude for one scanline from the Doll image pair that are shown in plot (a). Fig. 5.8 (b) shows the brightness that correspond to these scanlines from the left and right images. In order to estimate the disparity for the current pixel c , firstly, phase values with low corresponding magnitude will be filtered out using the unreliable phase constraint, Eq. 5.7, since these regions with low magnitude yield unstable phase values. Then use the remaining left and right phases to fit phase-differences. In order to do that, we will restrict our attention to those regions in which have visible magnitude values by finding peaks (singularity points) in the magnitude.

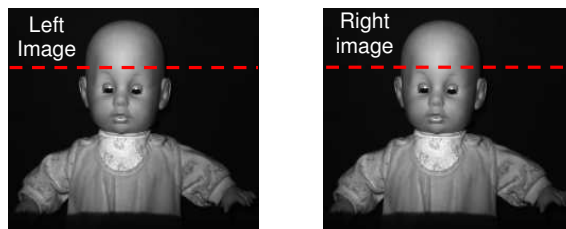
After that, for a given list of candidate disparity, d , (i.e. the disparity rang). Unlike the direct phase method, which make explicit use of phase difference as disparity, by comparing the phase values at the same-numbered column. we use columns that are actually expected to correspond, i.e. incorporate the candidate disparity in the measured-phase difference $\Delta_{\phi_{meas}}$, Eq. 5.11.

For more accuracy, we also should manipulate the complementary problem of *phase-wraparound*, wherever we can only measure the phase difference modulo 2π . I.e. a given filter will only be able to estimate disparity less than the wavelength specified by its tuning frequency [3] [68]. For more details on how we can address this problem, we enumerate the given list of the possible disparity range d , then for each candidate disparity $disp$, where, $disp \in [min(d), max(d)]$ we compute the "ideal-phase difference curve", ($\Delta_{\phi_{ideal}}$), from the wellknown relation between the phase-difference and the disparity, ($disp = \frac{\Delta_{\phi_{ideal}}}{2\pi}\lambda$) the ideal-phase difference given as;

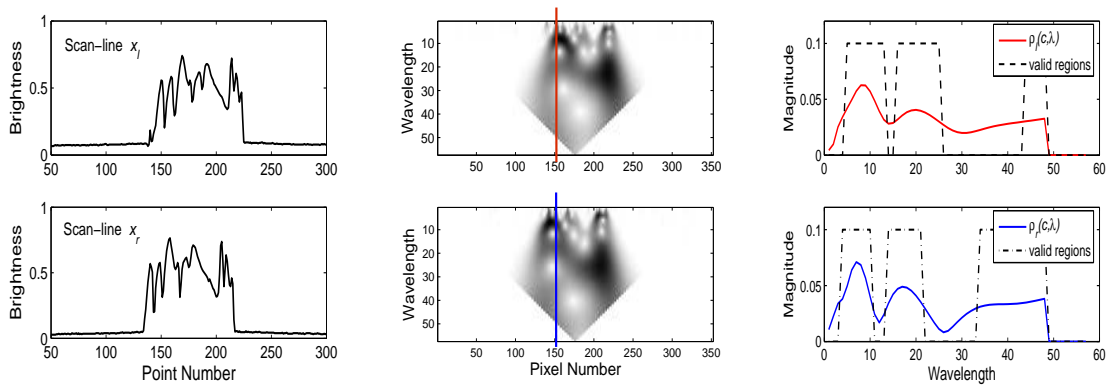
$$\Delta_{\phi_{ideal}} = 2\pi \frac{disp}{\lambda} \quad (5.8)$$

In the same time, the "measured-phase difference", $\Delta_{\phi_{meas}}$, defined as;

$$\Delta_{\phi_{meas}} = \phi_l(c, \lambda) - \phi_r(c, \lambda) \quad (5.9)$$



(a) Left and right stereo images, dashed line marked the scanline (row No. 75) to analyze



(b) The brightness shape of the scanline (c) Scalogram-magnitude of the scanline (d) Magnitude column No. 150

Figure 5.8: Sample usable magnitude that correspond to the stable phase values at pixel number (75,150) from the pair images. only the values inside the dotted rectangles are used in the computation.

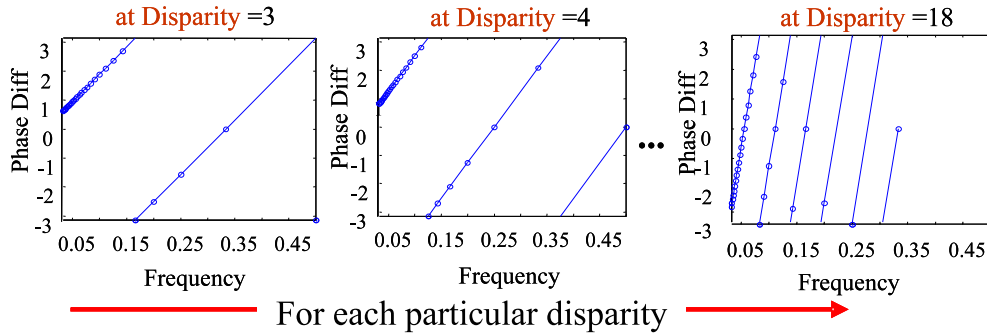


Figure 5.9: Ideal phase difference as a function of candidate disparity illustrate the phase wraparound problem. Each graph has 30 sample points spaced linearly along the x-axis.

This Equation measured phase-difference but, may be not equal the ideal-phase difference, recall the example in Fig. 5.2, since a wavelength $\lambda = 20$. Therefore, disparities of 1, 21, 41, ... are appear equivalent, as not expected. i.e. $\Delta x(1) = \Delta x(i\lambda + 1)$, $i = 1, 2, \dots, etc$. Therefore;

$$\Delta_{\phi_{meas}} = |\Delta_{\phi_{ideal}}|_{2\pi} \quad (5.10)$$

So for correcting the phase angle radian, one must add multiples of $\pm 2\pi$, then $\Delta_{\phi_{ideal}} = \Delta_{\phi_{meas}} + k2\pi$, which mean that the measured phase difference is only part of the results. For computing the exact disparity for each filter the additional parameter k must be known.

Unfortunately, there is no way to measure the k without knowing the ideal disparity, many methods either assume $k = 0$ or arrange the processing so that k is assumed to be known. Other direct phase methods [30] [4] address the phase wraparound problem by using a coarse-to-fine approach. Instead of that, we overcome this problem by add the given candidate disparity in the measured-phase difference computation and find the smallest correlation between the measured and ideal phase difference. Practically, we look at how the wraparound issue will arise, Fig. 5.9 shows several plots of ideal phase difference as a function of disparity range. As a candidate disparity increases, the number of times that phase value is expected to wraparound increases.

Therefore, Eq. 5.9 can be written as:

$$\Delta_{\phi_{meas}}(c, \lambda, disp) = \phi_l(c, \lambda) - \phi_r(c + disp, \lambda) \quad (5.11)$$

Now the task is turned to finding the disparity whose ideal-phase difference best match with that was measured. By this way we do not need to compute the appropriate value for k at each filter. Finally, an evaluation function, EF , compute a

quantitative agreement between these two set of phase differences.

$$EF = \frac{1}{|\mathcal{U}|} \sum_{\lambda \in \mathcal{U}} \rho_l(c, \lambda) ADM(\Delta_{\phi_{ideal}}, \Delta_{\phi_{meas}}) \quad (5.12)$$

where ADM (the absolute difference module) is the smallest difference between the ideal-phase and the measured phase, such that;

$$ADM(a, b) = \min_{k \in [-1, 0, 1]} ||a|_{2\pi} - |b|_{2\pi} + k2\pi| \quad (5.13)$$

where the \mathcal{U} is the set of wavelengths whose output are considered reliable. This evaluation function provides the most accurate results, where the smallest value indicates the best matches and the wraparound problem is addressed by finding the smallest ADM between $(\Delta_{\phi_{meas}})$ and $(\Delta_{\phi_{ideal}})$. Then we select the estimating disparity that exhibits minimum error as the result for the pixel.

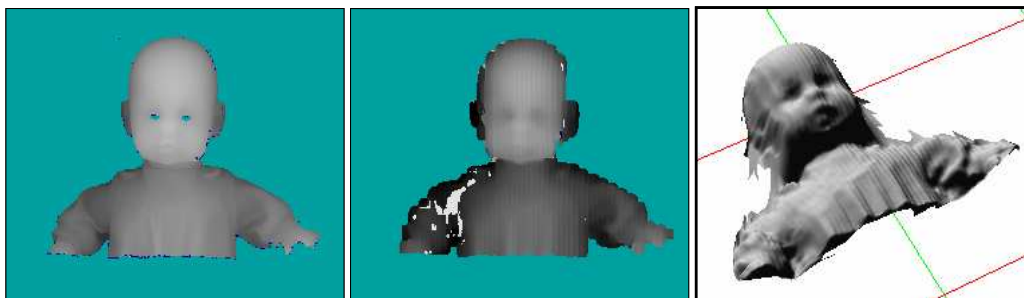
Fig.5.10 shows the computation results by the proposed phase-difference technique for the stereo pair image "Doll images", with disparity range (0...15), number of wavelengths $m = 4$ and number of sigma $\frac{1}{\sigma_f} = 6$. Fig.5.10(a) shows the true disparity map in the left and the computed disparity in the middle. Depth is coded in the disparity map as brightness, with brighter map areas estimated as nearer, darker areas estimated further away from the viewer. Some outliers areas also appears in white points. The disparity map converted to a 3-D point cloud then 3-D surface reconstructed as shown in the right plot, this is based on the known camera parameters. Fig. 5.10 (b) is a disparity space computed at row 75. Bottom plot Fig. 5.10 (c) is a representative scanline number 75 from a computed disparity map against the correspond one from true disparity map. The accuracy and the smoothness of the computed result are observed from the comparison, that is because the proposed method optimally captures both local orientation and frequency information from the input image.

5.3 Evaluation Methodology

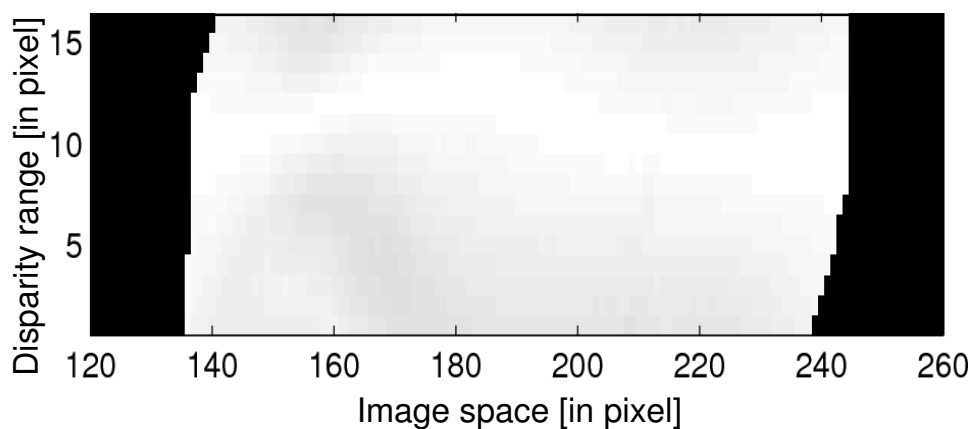
In order to evaluate the performance of our proposed algorithm and the effects of varying some of its parameters, we need a quantitative way to estimate the quality of the computed results.

First, we evaluate the results based on Scharstein and Szeliski data set [35] to have a comparison with various well-known stereo matching algorithms which provided at "cat.middlebury.edu/stereo" [1]. Table 5.2 summarizes the attributes of the four stereo pairs; dimensions, disparity rang, scaling and borders.

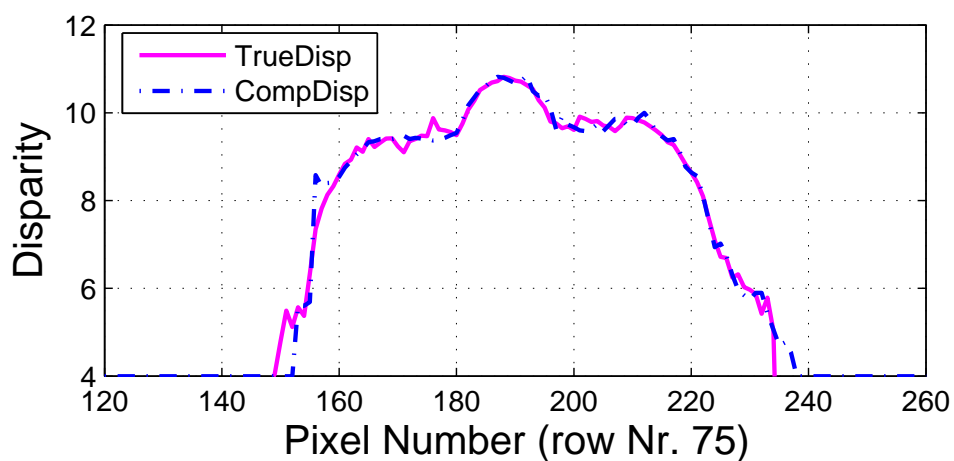
Two different statistics have been defined to measure the quality of results, that is based on known ground truth disparity maps:



(a) Ground truth (reference) disparity map (left), Computed disparity map (middle), Visualization of 3-D surface reconstruction (right)



(b) Disparity space computed for a line profile (row Nr. 75) from the Doll image pair



(c) Scanlines comparison from the computed resulted (CompDisp in a dash-dot line) and the true disparity (TrueDisp in a solid line)

Figure 5.10: Experimental results of the Doll image pair with disparity range (0...15). The accuracy and the smoothness of the computed result are observed from the comparison, that is because the proposed method is optimally captures both local orientation and frequency information from the input image.

- Percentage of bad matching pixels, (BadM), between the computed disparity, D_c , and the ground truth disparity, D_t ;

$$BadM = \frac{1}{N} \sum_{(x,y)} |D_c(x,y) - D_t(x,y)| > \delta_{error} \quad (5.14)$$

where δ_{error} is a disparity error tolerance. For the experiment in this work we use $\delta_{error} = 1.0$ since this coincides with Scharstein et al. [35], N is the total number of pixels.

- Root-Mean-Squared error (RMSE),

$$RMSE = \left(\frac{1}{N} \sum_{(x,y)} |D_c(x,y) - D_t(x,y)|^2 \right)^{\frac{1}{2}} \quad (5.15)$$

the mean square error is one of the most commonly used measures of success of numeric computation. As well as the root mean squared error gives the error value the same dimensionality as the actual and computed values.

The peak signal-to-noise ratio, often abbreviated PSNR, is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale.

The PSNR is most commonly used as a measure of quality of reconstruction. It is most easily defined via the root mean squared error (RMSE) which for two images D_c and D_t where one of the images is considered a true image and the other is the computed disparity.

The PSNR is defined as:

$$PSNR = 10 \cdot \log_{20} \left(\frac{\max_{pixel}}{RMSE} \right) \quad (5.16)$$

Here, MAX_{pixel} is the maximum pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255. More generally, when samples are represented using B bits per sample, maximum possible value of MAX_{pixel} is $(2^B - 1)$.

In order to obtain the algorithm ranking, the four different images that are described in Table 5.2 with their ground truth are used. The statistic based on the percentage of bad pixels is computed, furthermore to compute these statistic over whole image the computation also focus on three different kinds of regions: all pixels in non-occluded regions ($B_{\tilde{0}}$), all pixels in half-occluded regions (B_{ho}) and all pixels near-occluded regions i.e. near discontinuities (B_{no}). The numbers represent the percentage of bad pixels whose absolute disparity error greater than a threshold δ_{error} . The first measure, for example, defined as:

$$B_{\tilde{0}} = \frac{1}{|\tilde{0}|} \sum_{(x,y) \in \tilde{0}} (|D_c(x,y) - D_t(x,y)| > \delta_{error}) \quad (5.17)$$

Table 5.2: Attributes of four stereo pairs

Image pair	size	Disparity range	Scale	Border
Tsukuba	384×288	0..15	16	18
Venus	434×383	0..19	8	10
Teddy	450×375	0..59	4	0
Cones	450×375	0..59	4	0

similarly, B_{ho} and B_{no} are defined for the half-occluded and near-occluded regions. Fig.5.16, Table 5.3 show the application of there evaluations.

Second, we provide other three kind of error measures to evaluate the performance of the phase difference based algorithms;

- The absolute disparity error map (ADE):

$$ADE = |D_c(x, y) - D_t(x, y)| \quad (5.18)$$

- The mean relative error (MRE) defined as:

$$MRE = \frac{|D_c(x, y) - D_t(x, y)|}{|D_t(x, y)|} \quad (5.19)$$

- The percentage relative error (PRE):

$$PRE = 100 \cdot \frac{|D_c(x, y) - D_t(x, y)|}{|D_t(x, y)|} \quad (5.20)$$

In order to further simplify the comparison, it would be useful to express the whole error in the images by scalars. Therefore, we will refer to the mean value of these errors on a set of samples:

1. The average error (μ) which defined as the normalized some of absolute value of the difference between ground truth and computed disparity map at the nonsingular points. The points marked as unreliable, are simply discarded and are not taken in account in the computation. 5.2.2),

$$\mu = \frac{1}{N} \sum_{(x,y)} |D_c(x, y) - D_t(x, y)| \quad (5.21)$$

It is nearly equal to Eq. 5.14 but here x, y run over the nonsingular points.

2. The mean percentage relative error (PRE) defined as;

$$PRE = 100 \cdot \frac{1}{N} \sum_{x,y} \frac{|D_c(x, y) - D_t(x, y)|}{|D_t(x, y)|} \quad (5.22)$$

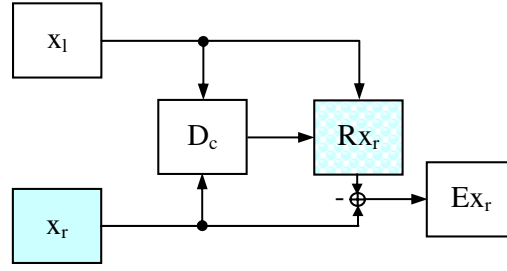


Figure 5.11: Principle of the quality measurement for disparity map without ground truth.

3. The deviation of the average error ($DevAE$);

$$DevAE = \frac{1}{N} \sum_{x,y} |\mu - |D_c(x, y) - D_t(x, y)|| \quad (5.23)$$

Third, we describe in this part an evaluation method in case of the the ground truth disparity map is not available. As the disparity map is obtained from a stereo pair images, it can reconstruct the right image by only using the left image and the computed disparity map. Considering the fact that the disparity map D_c contain the spatial shift between corresponding pixels in the stereo images. Therefore, the reconstruction of the right image defined as;

$$RI_r(x, y - D_c(x, y)) = I_l(x, y) \quad (5.24)$$

where x, y denote to the row and the column indices of the image. The difference between the original image I_r and the reconstruction RI_r yield the error image EI_r ;

$$EI_r(x, y) = I_r(x, y) - RI_r(x, y) \quad (5.25)$$

In practice, it can also express the whole error image in one scaler, by summation of all pixels and a consequent normalization of the image size yield the error criterion ξ ;

$$\xi = \frac{\sum_{x,y} |EI_r(x, y)|}{N \cdot N \cdot Rec} \quad (5.26)$$

where the parameter Rec represents the resolution of the image ($Rec = 2^8 = 256$) gray scale value.

Finally, percentage of both total *really wrong pixels* (\mathfrak{R}) and *corrected pixels* (η) are provided by counting the pixels whose values are actually error. The tests have been performed using a lot of stereo pairs either from real world or constructed.

5.4 Experimental results

In this section, we present the experimental results that are obtained from several stereo pairs and the quantitative evaluations as we mentioned in the evaluation methodology. At the beginning of our experiment, we analyze the performance of the phase-difference-based technique in relation with the choice of filter's parameters. After that the performance improvement is produced by the adaptive filter using different types of stereo images. Then a computed disparity is compared against ground truth disparity and also compared with the existence algorithms.

5.4.1 Tuning Parameters

The response of scalogram is characterized by the Gabor filters parameters: measures the width of local envelope (σ_f , m) and the tuning frequency (ω). The relative magnitude of these parameters determines how many oscillations of wavelength λ take place inside a region in which the filter is significantly different from zero.

Firstly, we discuss the properties of using a single filter on disparity estimates. In order to simplify the analysis, a one dimensional stereo signal composed of 256 points is created (e.g. the spike signal), wherever the left image is created by shifting the first spike by 1 pixel, the second by four pixels and the third by 2 pixels. Fig. 5.12 shows the effects of varying frequencies $\omega = \frac{1}{10}, \frac{1}{30}$, and $\frac{1}{50}$ on the performance of the computed disparity, the other two parameters have a constant values; $m = 4$ and $\sigma_f = 1/6$, that means we discuss the properties of a single filter.

The left graph in Fig. 5.12 (a) shows the real and imaginary representation (in top plot). The scalogram -magnitude and scalogram-phase are shown in the middle and bottom plots.

The right graph in Fig. 5.12 (a) illustrates the stereo pair images (left and right spike signals) in dash-dot line and in solid line respectively. Their correspond unwrap phases (s.phasL and s.phasR) are shown also in dash-dot and solid lines.

The computed disparity by our phase-difference method is shown in a dashed line. We can see the same plots in each case with a variety frequency where $\omega = \frac{1}{30}$ in Fig. 5.12 (b), and $\omega = \frac{1}{50}$ in Fig. 5.12 (c).

As shown in the same figure, by decreasing the frequency values, we can observe that the range area in scalogram-magnitude and scalogram-phase which differ from zero are decreased. See middle and bottom plots, S.mag and S.phase, in each case. As a result, the disparity will be only computed in these regions. That is because at a high frequency the filter width will be quite small, thus ensuring that only the nearest pixels will be used to compute filter outputs. At the low frequency the filter width will be much wider, and will therefore require more number of pixels that must be sampled. Therefore, the choice of the frequency should consider the fact that the wavelength $\lambda = 1/w$ has to be at least twice as large as the expected maximum disparity [31], [69].

Fig. 5.13 illustrates the effect of varying the fraction of standard deviation $\sigma_f = \frac{1}{4}, \frac{1}{6}$ and $\frac{1}{8}$ while the frequency keep fixed at $\omega = \frac{1}{10}$ and the number of periods $m = 4$.

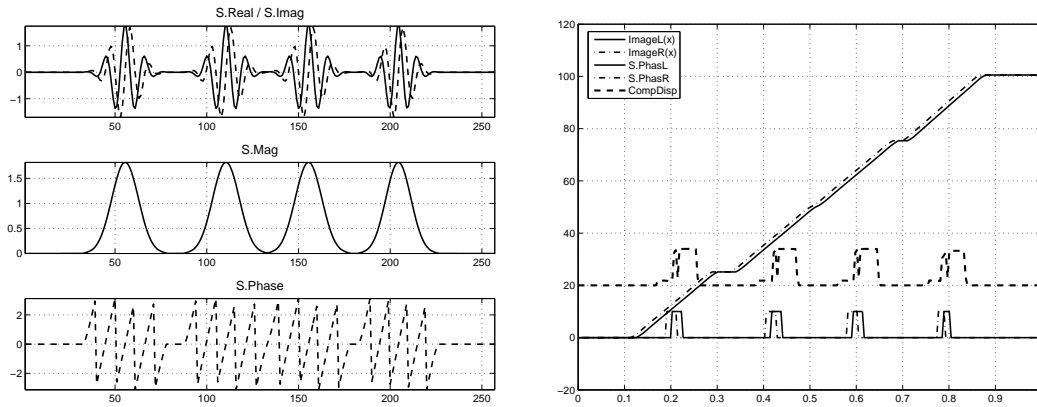
As we can see in the graphs, the disparity estimates are effecting by extent of the Gaussian window. In each case of Fig. 5.13 (a), (b) and (c) we can notice that the magnitude (S.Mag) and the phase (S.Phase) behaviors in the x-axis region (100 ... 150) in each case at varying values of σ_f and observe their effected in the computed disparity. Seeing that, the phase is a quasi-piecewise linear function, the transition between quasi linear regions occur when the magnitude is very small, $\rho \simeq 0$, observe the magnitude and the phase graph in each case. This transition is proportional to the width of Gaussian envelop.

Similarly, Fig. 5.14 demonstrates the effect of varying number of wavelengths m on the disparity estimates. All of the filters in the figure have constant frequency $\omega = 1/10$ and $\sigma_f = 1/6$, but the extra parameter m varies with values; 6, 10 and 14. From the graphs, we can observe the amount of lost information due to the change of sampling the signal. Therefore the performance of the disparity is changed.

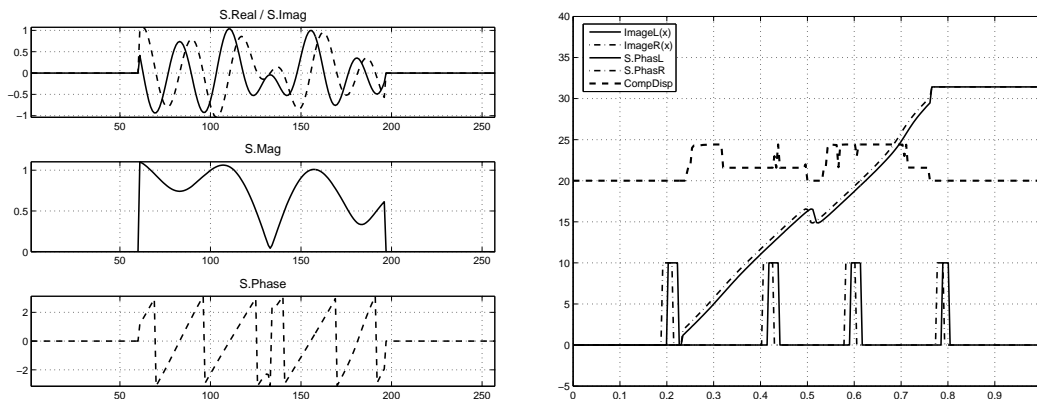
From these analysis, the performance is decisively improved if wavelength and the filter width are changed concomitantly, i.e. by keeping fixed relation between wavelength and width, since the smoothness of the phase function is improved by increasing the the width of fixed frequency. But large filter can cause a loss of resolution and a consequent error increase. By experiments, a good compromise is choosing $\sigma \simeq 1/2\omega$, in this article we chose the value of $m = 4$ and $\sigma_f = 1/8$, that allows lots of data and many wavelength to be used in the computation. Once the relation between width and wavelength is chosen the filter's performance becomes a one parameter function (the tuning frequency). In the real image, which have unfixed frequency in its scanlines it will be more convenient to use many filters (i.e. the local spatial frequency models) that make use of local frequency content of an image.

Now, we analyze the performance of the phase-difference based technique for disparity estimation by using an adaptive kernel window that shrinks and grows as the tuning frequency changes.

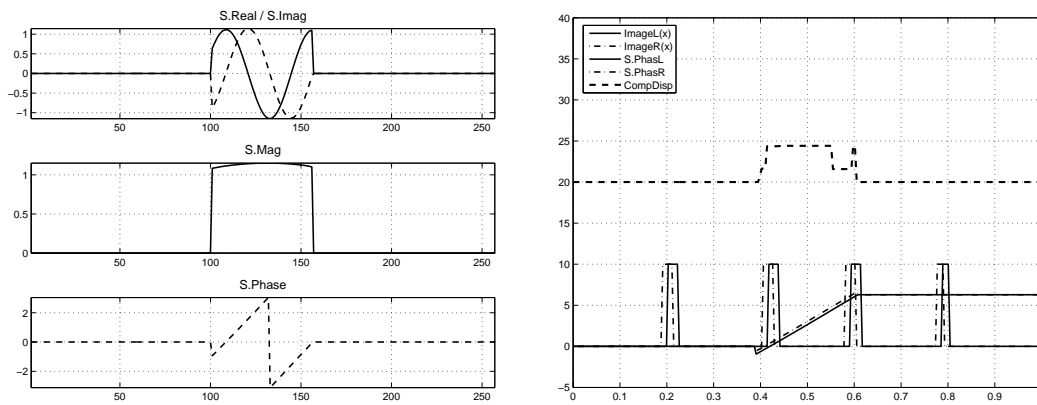
Fig. 5.15 shows in each case; (a), (b) and (c) the intensity representation of a line profiles from a real stereo pair images "Tsukuba" (left column, Top), scalogram representation (phase, magnitude) for the left and right scanlines shown in the middle and bottom left column in each case. The upper plot in right column is the disparity values for each pixel with the candidate disparity (disparity rang, 0 ... 16), lower plot is a graph of representative scanline from a computed disparity map (in solid line) compared to the ground truth disparity (in dashed line). Fig. 5.15 (c) shows in a comparison plot the improvement of the result, where the computed disparity (CompDisp) and ground truth (TrueDisp) are nearly closed to each other.



(a) Choice of filter Parameters at $\omega = 1/10$, $\sigma_f = 1/6$ and $m = 4$

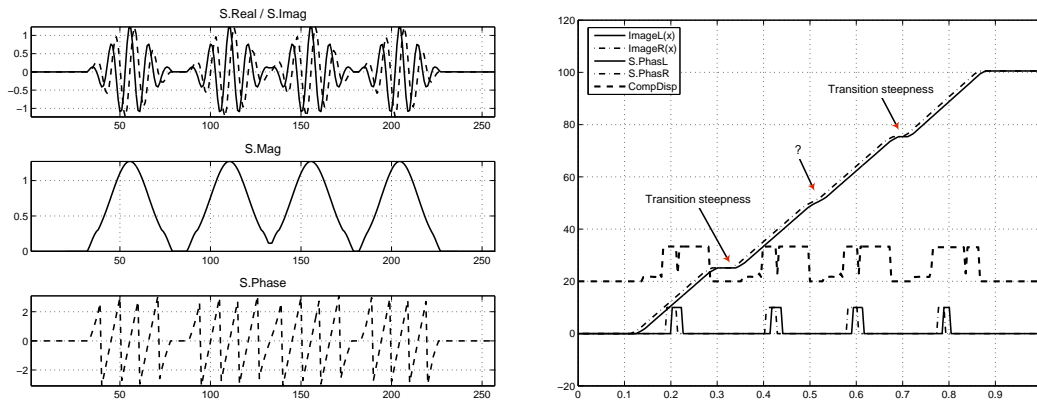


(b) Choice of filter Parameters at $\omega = 1/30$, $\sigma_f = 1/6$ and $m = 4$

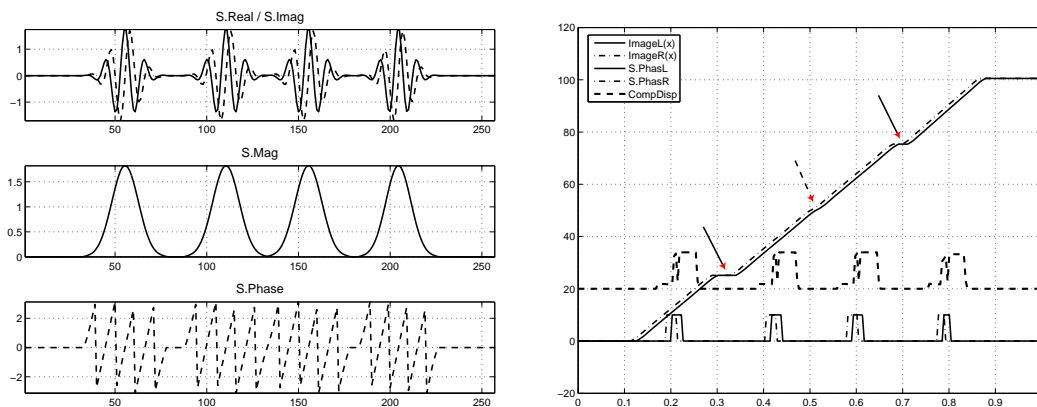


(c) Choice of filter Parameters at $\omega = 1/50$, $\sigma_f = 1/6$ and $m = 4$

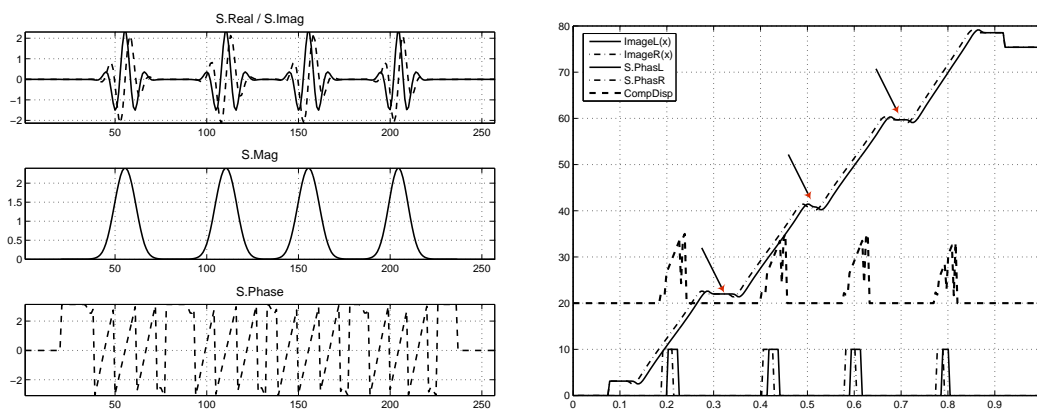
Figure 5.12: The choice of frequency. The computed disparity at $\omega = 1/10, 1/30, 1/50$, while two other parameters (m, σ_f) are kept fixed at $m = 4, \sigma_f = 1/6$, disparity range $[0..15]$. the stereo images composed of $N=256$ points (as pike, like one dimensional stereo signal). The figures are translated on the Y axis for visualization.



(a) Choice of filter Parameters at $\omega = 1/10$, $\sigma_f = 1/4$ and $m = 4$



(b) Choice of filter Parameters at $\omega = 1/10$, $\sigma_f = 1/6$ and $m = 4$



(c) Choice of filter Parameters at $\omega = 1/10$, $\sigma_f = 1/8$ and $m = 4$

Figure 5.13: Effect of σ_f . The transition steepness is marked by the arrows.

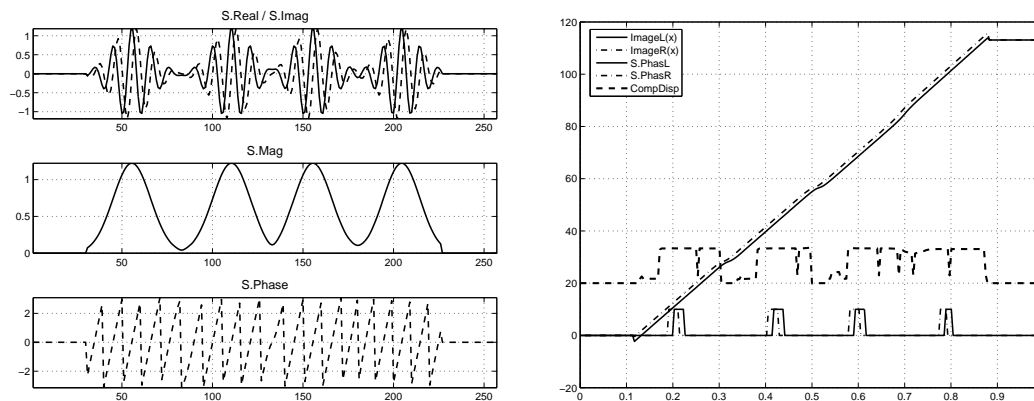
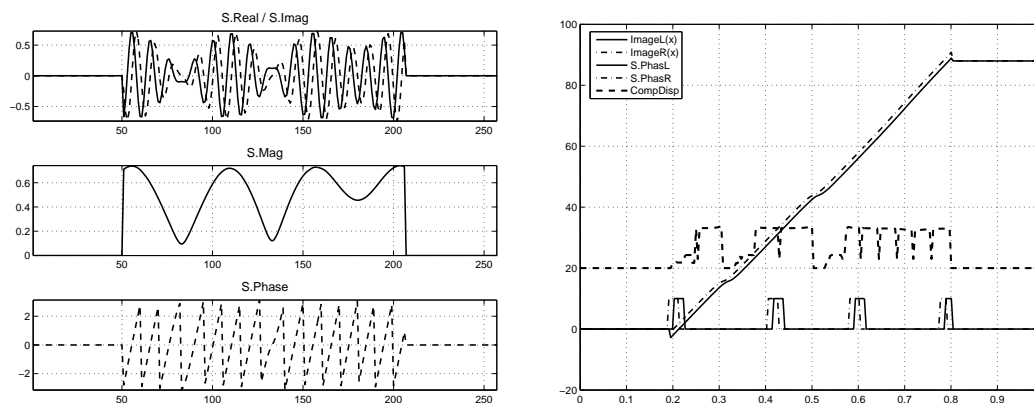
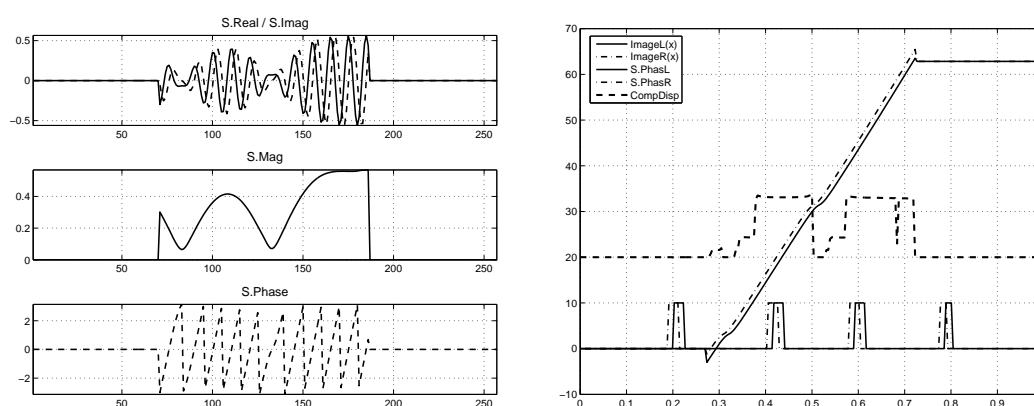
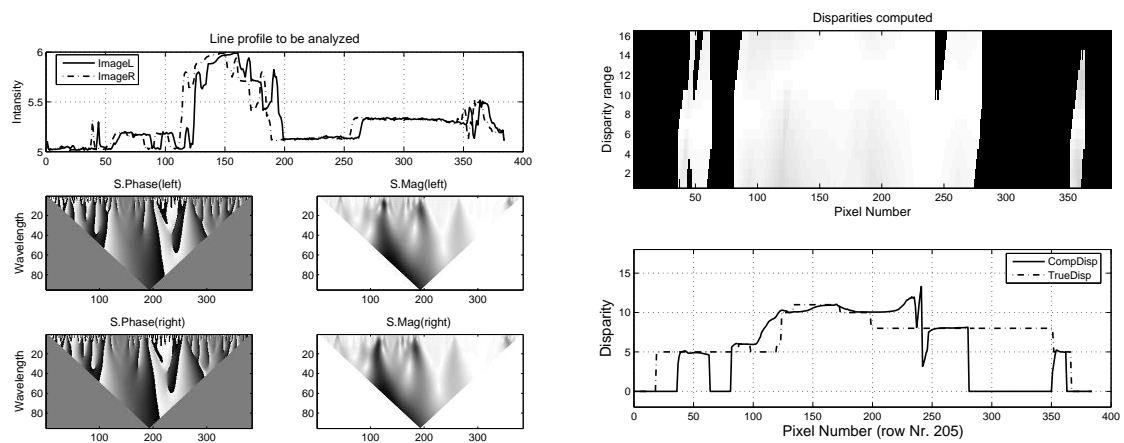
(a) Choice of filter Parameters at $\omega = 1/10$, $\sigma_f = 1/6$ and $m = 6$ (b) Choice of filter Parameters at $\omega = 1/10$, $\sigma_f = 1/6$ and $m = 10$ (c) Choice of filter Parameters at $\omega = 1/10$, $\sigma_f = 1/6$ and $m = 14$

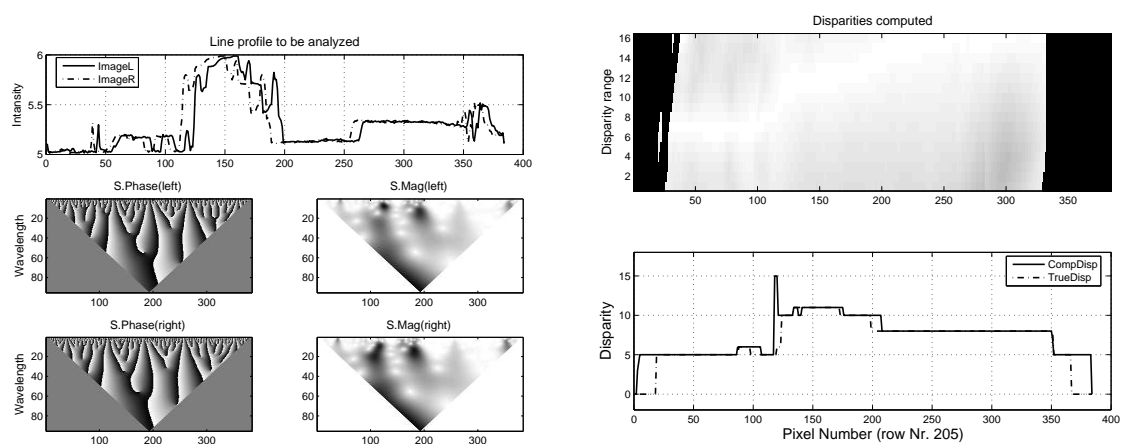
Figure 5.14: Effect of a varying number of wavelengths m on the disparity estimates. All of the filters in the figure have constant frequency $\omega = 1/10$ and $\sigma_f = 1/6$.



(a) Left and right "Tsukuba" stereo image and its ground truth map from data set [1]



(b) Computed disparity, $\sigma_f = 1/10$ and $m = 4$



(c) Computed disparity, $\sigma_f = 1/6$ and $m = 4$

Figure 5.15: Line profile of computed disparity from our adaptive filter compared with the ground truth in different values of σ_f .

Table 5.3: Results of evaluation methodology

	Tsukuba	Venus	Teddy	Cones
BadM	0.255	0.345	3.14	3.16
RMS	3.26	1.94	8.89	9.17
μ	0.323	0.605	4.49	3.47
PRE	5.50	8.76	11.0	7.71
DevAE	0.533	0.599	4.89	4.91
η	83.4	84.5	82.0	81.8

5.4.2 Ground-truth Evaluation

This section provides experimental results obtained with the proposed "phase-difference" algorithm on a standard set of stereo pairs (namely Tsukuba, Venus, Teddy, and Cones) with available ground truth. The stereo pairs and the ground truth are available at [1]. We used disparity ranges, scales and border values as mentioned in Tab.5.2. The data set of the Middlebury stereo web side [1] are shown in Fig. 5.16. The left images of each pair shows in Fig. 5.16(first row). The ground truth disparity maps are shown in (second row), the computed disparity in the (third row), Error maps are shown in the (fourth row), where the outlier pixels are assigned in black areas and qualify as the most difficult areas for the algorithm. The bottom row is the left occluded areas that are excluded and shown in the white regions. The total matching error is calculated as the mean percentage of outliers (disparity error > 0.75) over all data sets.

As we mentioned in the evaluation methodology section 5.3, the computed disparity maps are then compared against the ground truth by computing the percentage of wrong pixels. A pixel is judged to be erroneous, if its absolute deviation from the ground truth is larger than 0.75 unit. Furthermore, Table 5.3 illustrates the statistics results for all data sets in a quantitative measures to evaluate the quality of the computed results. The percentage of bad matching pixels, Root mean square error, average error for the nonsingular points, the mean percentage relative error, the deviation of the average error and the percentage number of correct pixels η are reported.

5.4.3 Comparison With Other Algorithms

Table 5.4 compares the proposed algorithm with 31 other existing algorithms. The percentage of bad pixels, (the pixels which deviate is more than 0.75 unit from the true disparity and labeled as "*bad pixels*"), in the entire image, in the half-occluded regions and near occluded regions are used to compare the results of various algorithms for example; Improve SubPix [70], C-SemiGlo [71], GenModel [36], DoubleBP [74], AdaptWeig [75], STICA [78], SSD+MF [35]. The table shows the percentage of

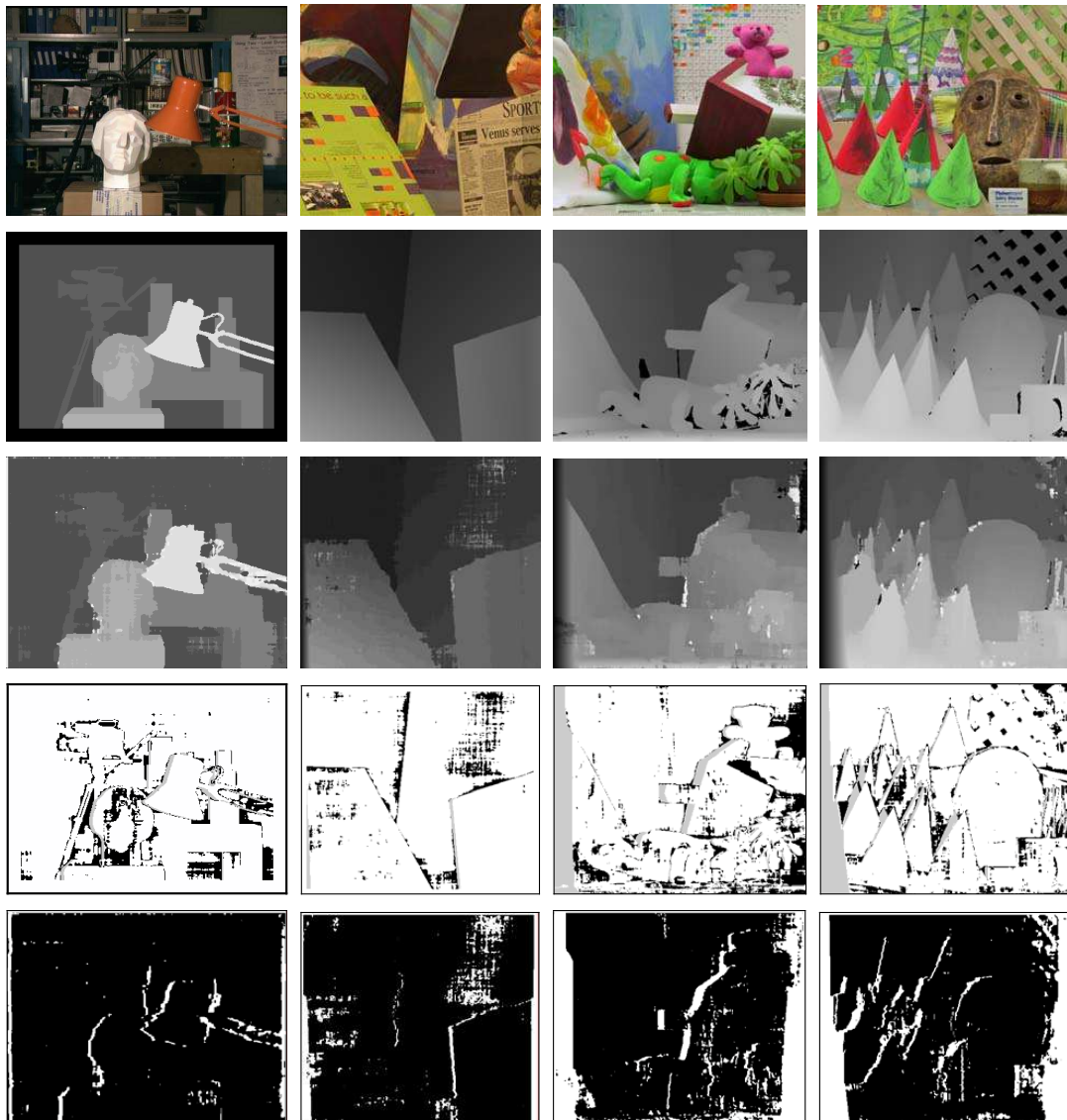


Figure 5.16: Resulting disparity maps from the Middlebury stereo data set. (top row) Left image; (second row) Ground truth [1]; (third row) Disparity maps generated from our method; (forth row) Error map contain bad pixels (in black), correct pixels (in white) and excluded occlusions (in gray);(bottom row) is the occluded region assigned in the white areas.

Table 5.4: Performance comparison from the Middle bury stereo vision page [1]. Error percentages and rank in each column is shown. Error threshold = 0.75. The table show only the ten top algorithms and the two latest algorithms in the evaluation

Algorithm	Tsukuba			Venus			Teddy			Cones		
	$B_{\tilde{O}}$	B_{ho}	B_{no}	$B_{\tilde{O}}$	B_{ho}	B_{no}	$B_{\tilde{O}}$	B_{ho}	B_{no}	$B_{\tilde{O}}$	B_{ho}	B_{no}
[70]	5.98 ₍₃₎	6.69 ₍₂₎	13.4 ₍₃₎	1.53 ₍₁₃₎	2.19 ₍₁₄₎	9.60 ₍₁₆₎	8.43 ₍₇₎	14.6 ₍₈₎	19.3 ₍₆₎	3.40 ₍₂₎	8.98 ₍₁₎	9.55 ₍₃₎
[71]	6.21 ₍₅₎	6.90 ₍₃₎	12.9 ₍₂₎	0.46 ₍₂₎	0.85 ₍₃₎	4.42 ₍₈₎	6.39 ₍₃₎	13.8 ₍₅₎	16.0 ₍₃₎	3.29 ₍₁₎	9.21 ₍₂₎	9.38 ₍₂₎
[72]	6.10 ₍₄₎	7.11 ₍₄₎	14.6 ₍₆₎	3.20 ₍₂₃₎	3.80 ₍₂₂₎	8.40 ₍₁₅₎	14.7 ₍₂₃₎	21.3 ₍₂₃₎	24.4 ₍₂₁₎	6.81 ₍₂₀₎	14.0 ₍₂₀₎	15.8 ₍₂₀₎
[73]	5.28 ₍₂₎	7.47 ₍₅₎	15.8 ₍₁₁₎	2.04 ₍₁₉₎	3.45 ₍₂₀₎	17.7 ₍₂₆₎	9.19 ₍₁₁₎	17.4 ₍₁₈₎	23.1 ₍₁₈₎	5.97 ₍₁₅₎	16.3 ₍₂₃₎	13.4 ₍₁₆₎
[73]	8.26 ₍₁₀₎	10.4 ₍₁₀₎	23.0 ₍₂₆₎	2.83 ₍₂₁₎	3.89 ₍₂₃₎	29.1 ₍₃₃₎	16.5 ₍₂₆₎	25.1 ₍₂₆₎	34.0 ₍₃₁₎	9.15 ₍₂₃₎	17.6 ₍₂₅₎	22.2 ₍₂₆₎
[74]	18.7 ₍₂₀₎	19.1 ₍₁₉₎	15.8 ₍₉₎	0.83 ₍₄₎	1.34 ₍₆₎	3.84 ₍₃₎	5.77 ₍₂₎	11.4 ₍₄₎	13.4 ₍₁₎	4.80 ₍₉₎	11.3 ₍₁₀₎	11.6 ₍₈₎
PhasBased	11.2 ₍₁₁₎	13.4 ₍₁₂₎	23.9 ₍₂₈₎	7.81 ₍₃₁₎	9.28 ₍₃₁₎	27.9 ₍₃₂₎	30.4 ₍₃₁₎	28.5 ₍₃₁₎	32.7 ₍₂₉₎	15.9 ₍₃₂₎	25.2 ₍₃₃₎	27.5 ₍₃₁₎
[75]	18.1 ₍₁₉₎	18.8 ₍₁₈₎	18.6 ₍₁₇₎	1.32 ₍₁₁₎	109. ₍₁₁₎	8.27 ₍₁₄₎	10.1 ₍₁₆₎	16.3 ₍₁₄₎	23.5 ₍₁₉₎	6.17 ₍₁₆₎	12.1 ₍₁₄₎	12.1 ₍₉₎
[76]	19.0 ₍₂₁₎	20.7 ₍₂₅₎	17.5 ₍₁₄₎	3.98 ₍₂₆₎	5.20 ₍₂₅₎	15.9 ₍₂₅₎	14.7 ₍₂₄₎	21.6 ₍₂₄₎	25.1 ₍₂₂₎	15.7 ₍₃₁₎	22.6 ₍₃₀₎	23.5 ₍₂₈₎
[77]	19.1 ₍₂₂₎	19.5 ₍₂₀₎	20.0 ₍₂₁₎	1.67 ₍₁₇₎	2.05 ₍₁₃₎	4.63 ₍₉₎	9.28 ₍₁₂₎	15.2 ₍₁₁₎	20.2 ₍₉₎	5.62 ₍₁₂₎	11.8 ₍₁₁₎	13.1 ₍₁₄₎
:	:	:	:	:	:	:	:	:	:	:	:	:
[78]	24.3 ₍₃₃₎	26.1 ₍₃₄₎	44.8 ₍₃₅₎	9.65 ₍₃₃₎	11.0 ₍₃₃₎	42.6 ₍₃₅₎	18.2 ₍₂₇₎	25.8 ₍₂₇₎	42.8 ₍₃₄₎	11.6 ₍₂₇₎	19.6 ₍₂₆₎	33.1 ₍₃₄₎
[35]	28.5 ₍₃₅₎	30.0 ₍₃₅₎	38.1 ₍₃₄₎	4.98 ₍₂₈₎	6.43 ₍₂₈₎	14.5 ₍₂₄₎	19.8 ₍₂₉₎	27.8 ₍₃₀₎	38.2 ₍₃₃₎	13.8 ₍₃₀₎	22.8 ₍₃₁₎	31.2 ₍₃₂₎

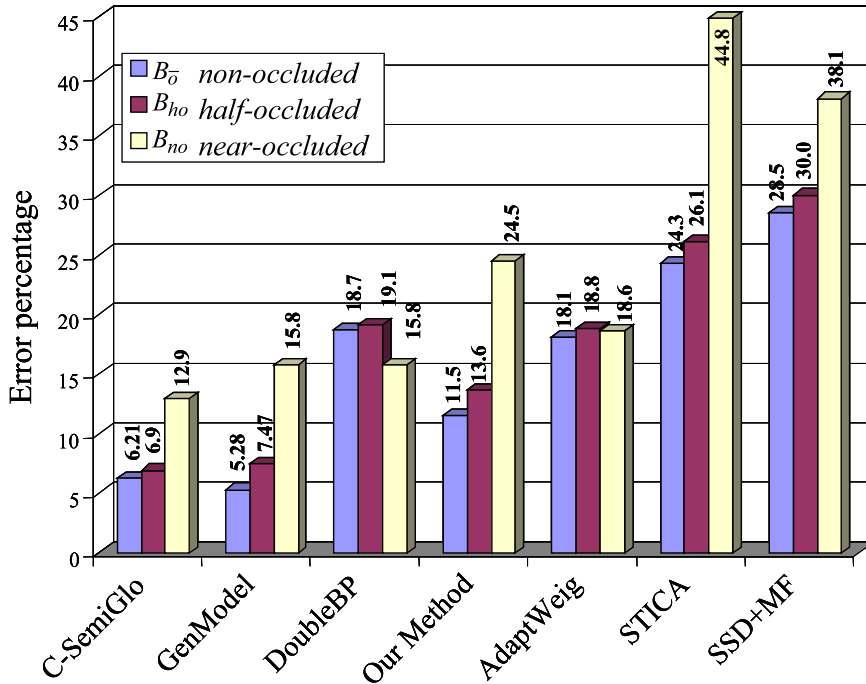
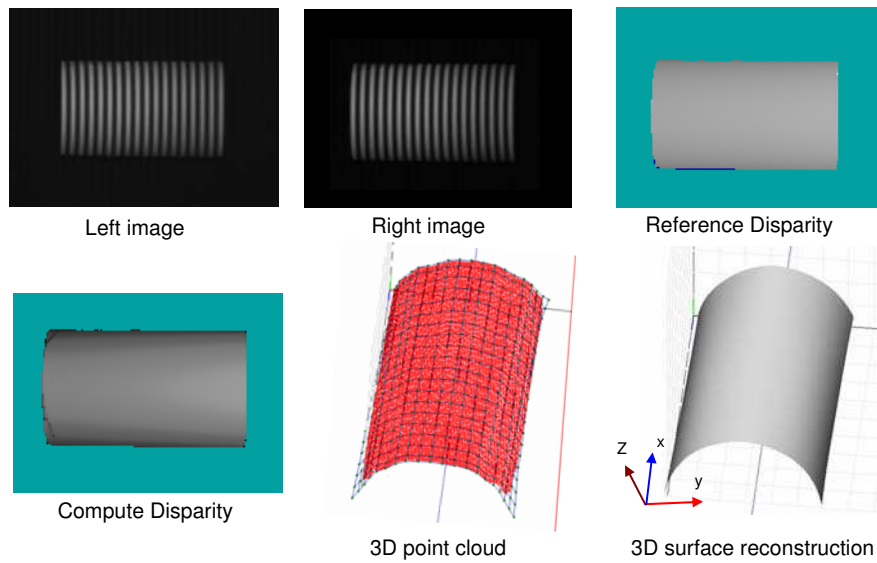


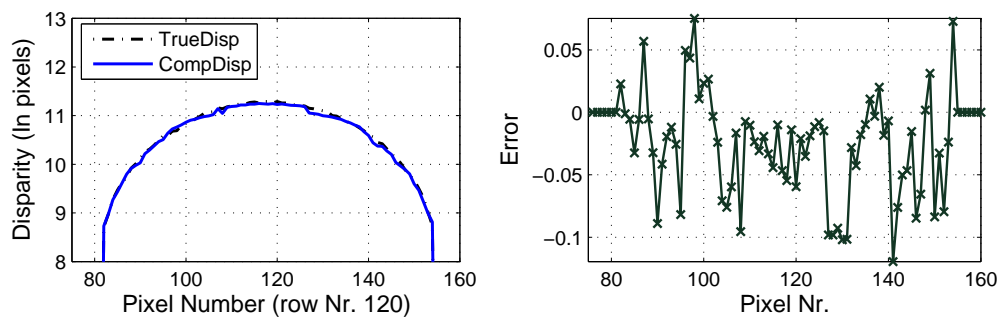
Figure 5.17: Percentage of bad pixels in $B_{\bar{o}}$, B_{ho} and B_{no} of various algorithms.

bad pixels, which are generated from our disparity maps Fig. 5.15 to the web site "http://vision.middlebury.edu/stereo" [1]. The ranks in each column are shown in brackets, as a subscript of the error percentage. Each algorithm is sorted according to its overall ranks. Fig. 5.17 shows the behavior of the percentage of bad pixels in: non-occluded region, half-occluded region and near-occluded (i.e. near-discontinuities) region of some various algorithms.

Fig. 5.18 shows another example, "cylinder view", the figure represents the computed disparity map, the 3-D point cloud and the 3-D surface reconstruction. A representative comparison of correspondence lines profile from reference disparity map (TrueDisp in dashed line) and the computed disparity map (CompDisp in solid line) is shown in Fig. 5.18 ((b)-left), the right plot show the amount of error in each pixel along the scan line, the maximum error is less than 0.2 pixel. From these results we can say that the disparity estimates are obtained with subpixel accuracy, without requiring explicit subpixel signal reconstruction.



(a) Stereo pair image "Cylinder" with its ground truth and the computed disparity maps



(b) lines profile comparison from the computed (CompDisp) and the reference disparity (TrueDisp)

Figure 5.18: 3-D surface reconstruction for cylinder stereo pair images.

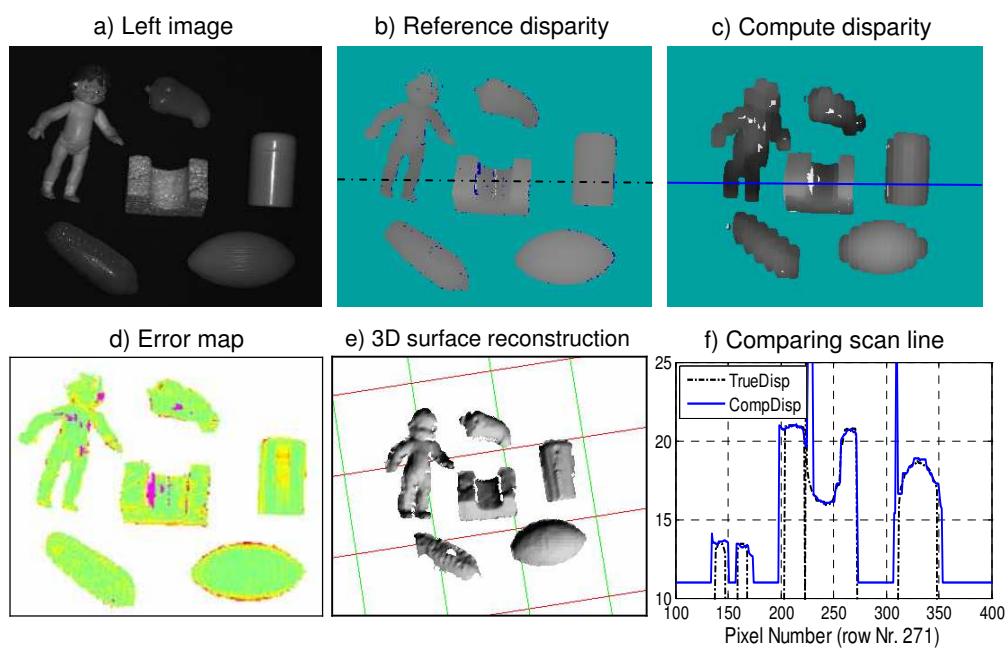


Figure 5.19: Multi object scene. (a) left image from the stereo pair, (b) reference disparity map, (c) computed disparity, (d) is the visualization of Error map where the correct pixels shown in gray-level from 32 to 64, the invalid pixels in blue and the bad matching pixels in magenta, (e) shows visualization of 3-D surface reconstruction, (f) lines profile comparison from the computed (CompDisp) and the reference disparity (TrueDisp).

Chapter 6

Correspondence with Slanted Surface: Foreshortening Effects

In this chapter, we describe the effect of perspective foreshortening in terms of local spatial frequency. We will develop this theory in steps to demonstrate several properties: the frequency shift between images of an oriented flat surface is constant. It is independent of the surface texture, and it can be expressed using only disparity range and surface angle.

The challenges for a slanted surface in the phase-based method arise in case of the instantaneous frequencies that measured by the same filter are different in a pair of corresponding image regions. Some authors; for instance [30] have smoothed over the effects by using the average of the two instantaneous frequencies, but these are coarse approximations that are not based on physical reality. In the following, we demonstrate a theory of modeling the physical effects of perspective distortion in both the world model and the stereo images.

In order to simplify the analysis, we assume that the only object in the 3-D world is a textured flat surface that is either parallel to the image plane, or rotated about the vertical axis by some angle θ , as shown in Fig.6.4. We further assume that the stereo cameras have parallel optical and vertical axes. Then, we restrict our attention to the effects of foreshortening in one-dimensional image scanlines, rather than completing two-dimensional images, since all disparities will be horizontal under this assumption. Fig. 6.1 shows an overhead view of foreshortening model and the effect of this foreshortening in the frequency domain. This transformation will be quantified precisely in the closed-form foreshortening factor developed in this section. We denote the parameters measuring distances in the world plane be capitalized, X_S , Z_L .

Now, we should looking for how the appearance of the object's surface texture changes between the perspective image planes, i.e. how the sampling rate varies between the two images. This can be described and simplified as follows: For each distance X_S along the world surface, we want to compare the projection segments x_L

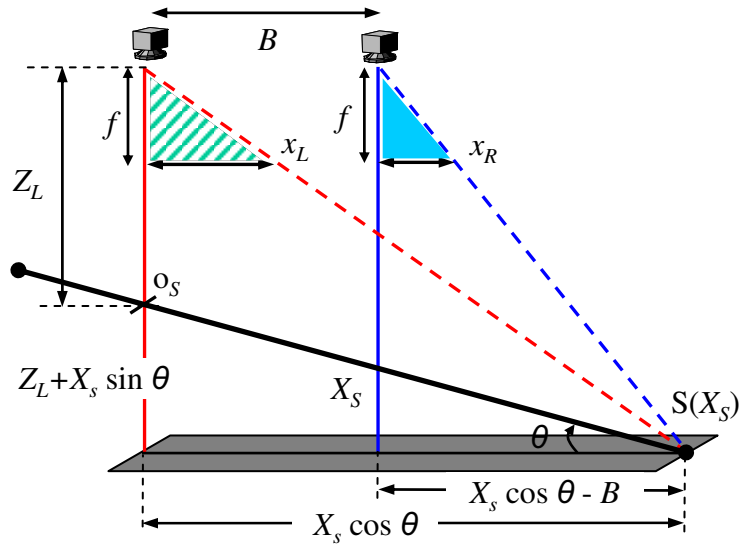
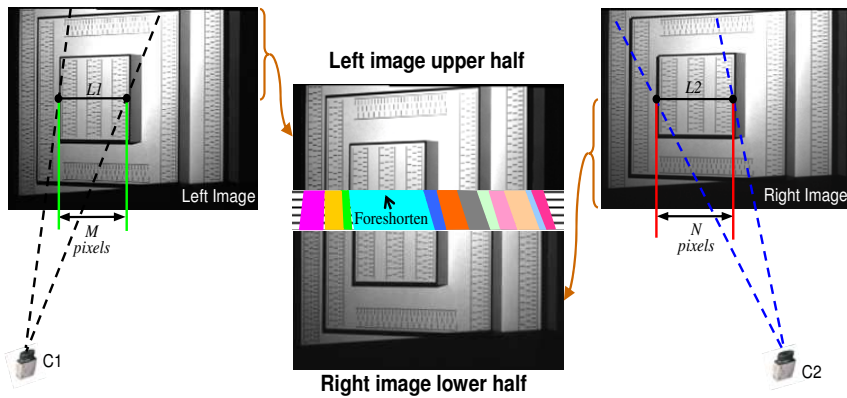
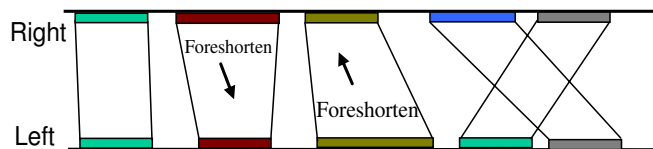


Figure 6.1: Foreshortening model. X_S is the distance from the original point, O_S , which exactly in front of the left camera to the point S on the surface. X_L and X_R are the left and right segment indices of the image for surface distance X_S ; the surface is viewed at a sharp angle θ .



(a) Left and right images where the corresponding projections upper-half from left image and lower-half from right image appears in different length



(b) Unequal correspondence intervals on the left and right scanlines should be correspond to each other

Figure 6.2: Unequal projection lengths for the object surface acquired in a slanted view.

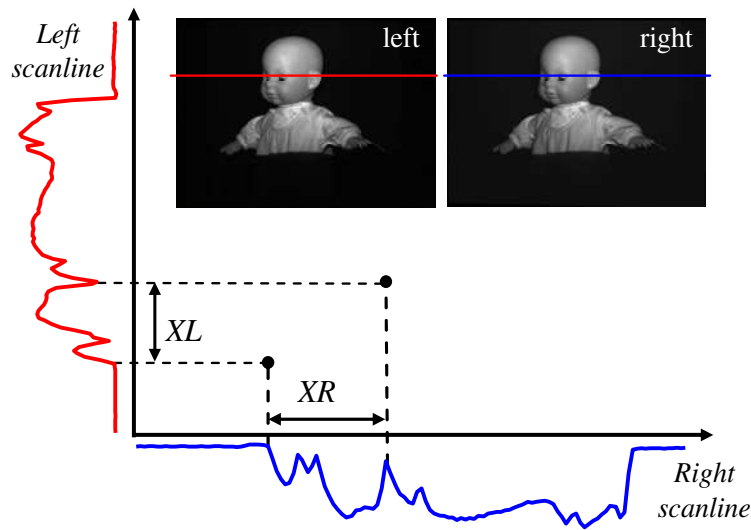


Figure 6.3: Stereo pair images show more compressed in one image due to perspective foreshortening. The projected segment X_L is shorter than X_R .

and x_R , from the left and the right images respectively, i.e. compare the left sampling rate ($\delta X_S / \delta x_L$) to the right sampling rate ($\delta X_S / \delta x_R$), consequently:

$$\Psi = \text{sampling ratio} = \frac{\delta X_S}{\delta x_L} \div \frac{\delta X_S}{\delta x_R} = \frac{\delta x_R}{\delta x_L} \quad (6.1)$$

This sampling ratio will be called the *foreshortening factor*, which is denoted as Ψ . As a result, this formula Eq.6.1 computes the sampling ratio in the image space, without having to explicitly model the distance X_S along the object. Unfortunately, this implies that it needs not only the left projected image x_L but also the correspondence projected x_R , which are the component of the disparity estimation. But the purpose of the work, as well as all methods that solve the stereo problem, is to estimate disparities values (i.e. search for each x_R in the right view that correspond to each x_L in the left view). So it would be better if the disparity and its derivative could be avoided in the calculation of the foreshortening factor. In the following two subsection 6.1 and 6.2, we will express the foreshortening factor without requiring the disparity value. The two issues of concern that will be considered are: First, relating disparity to the surface angle. Second, expressing the foreshortening factor uses the image parameters.

6.1 Relating Disparity to Surface Angle

The disparity is the difference between the left and right pixel indices. So, let us see how each of the left and right indices (x_L and x_R) relates to the surface angle θ .

Based on the foreshortening model which is presented in Fig. 6.1, we will focus our attention on the distance from the left camera to the surface point immediately in front of it, expressing other depths in terms of this value Z_L . The point varies across the world surface, which can lead to the changes of its projection on the image plane. The relation between the projection points in the image plane (*pixel index*) and the surface angle is calculated using similar triangles in Fig.6.1:

$$\frac{\text{pixel index}}{\text{focal length}} = \frac{X \text{ world coordinate}}{Z \text{ world coordinate}} \quad (6.2)$$

therefore, we obtain expressions for x_L and x_R from the similar triangles for the left and right camera geometry;

$$\frac{x_L}{f} = \frac{X_S \cos \theta}{Z_L + X_S \sin \theta}, \quad \frac{x_R}{f} = \frac{X_S \cos \theta - B}{Z_L + X_S \sin \theta} \quad (6.3)$$

Eq. 6.3 give us an expression for x_L and x_R in terms of the focal length f , baseline B , distance in front of the left camera Z_L , surface angle θ and location on the surface X_S . As a result, the equivalence expression for X_S from the left and right camera geometry is written as:

$$X_S = \frac{x_L Z_L}{f \cos \theta - x_L \sin \theta}, \text{ or } X_S = \frac{x_R Z_L + Bf}{f \cos \theta - x_R \sin \theta}$$

this equation represent projections of the same surface point X_S into two image plans. Therefore,

$$\frac{x_L Z_L}{f \cos \theta - x_L \sin \theta} = \frac{x_R Z_L + Bf}{f \cos \theta - x_R \sin \theta} \quad (6.4)$$

solving Eq. 6.4 for the right pixel index x_R gives:

$$x_R = x_L \left(1 + \frac{B}{Z_L} \tan \theta \right) - \frac{Bf}{Z_L} \quad (6.5)$$

Similarly, we can compute the left pixel index x_L . subsequently, to create the relationship between the disparity and the surface angle, recalling the equation that computes the disparity as the difference between the left and right indexes; $\text{disparity} = x_L - x_R$ and substituting by x_R , then the disparity can be written as:

$$\text{disparity} = \frac{B \cdot f}{Z_L} - x_L \frac{B}{Z_L} \tan \theta \quad (6.6)$$

This equation relates disparity to the scene parameters and does not depend on knowing the actual surface location. But it still requires knowledge of Z_L (distance to the surface point in front of the left camera).

6.2 Foreshortening-factor using Image Parameters

Eq. 6.6 relates disparity to the scan parameters, focal length f , baseline B . So we can observe when the surface is frontoplaner (i.e., $\theta = 0$). It is reduced to the familiar expression that relate disparity to depth; $disparity = \frac{f \cdot B}{Z_L}$, and for an arbitrary fixed angle θ the disparity varies linearly with respect to the image location x_L . As we can see in the experimental results, see the slanted flat surface in Fig.6.5. But Eq. 6.6 require knowledge of Z_L , to eliminate this restriction, let us return to the sampling ratio, Eq. 6.1, and reducing the derivative by substituting by x_R , Eq. 6.5, then this ratio (foreshortening-factor) is written as:

$$\begin{aligned}\Psi &= \frac{\delta \left(x_L \left(1 + \frac{B}{Z_L} \tan \theta \right) - \frac{Bf}{Z_L} \right)}{\delta x_L} \\ &= 1 + \frac{B}{Z_L} \tan \theta\end{aligned}\quad (6.7)$$

This formula gives the geometric form of the foreshortening factor. It tells us that for a given flat surface, the sampling ratio is constant over both images of the surface. In other words, the local spatial frequencies of the left and right images are related by a simple constant scale factor. We can get a feel for this by visually tracking the low magnitude phase singularities (white spots) between the two image scalograms in Fig. 6.4.

The result in Eq. 6.7 is useful for describing the form of foreshortening effect for a constant scale factor, but it would be useless in a stereo matching since it requires the knowledge of the depth Z_L . To eliminate the distance in front of the left camera Z_L , from Eq. 6.7, we return to Eq. 6.6 and reform it to:

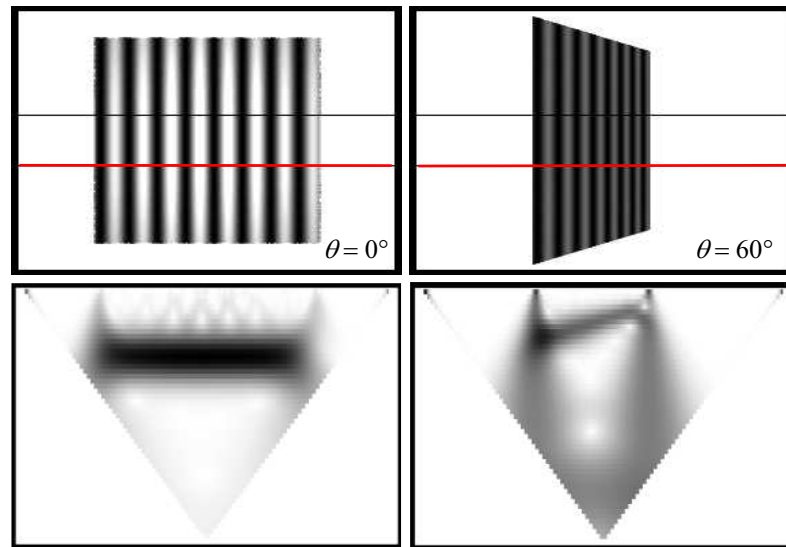
$$disparity = \frac{B}{Z_L} (f - x_L \tan \theta) \quad (6.8)$$

$$\frac{B}{Z_L} = \frac{disparity}{f - x_L \tan \theta} \quad (6.9)$$

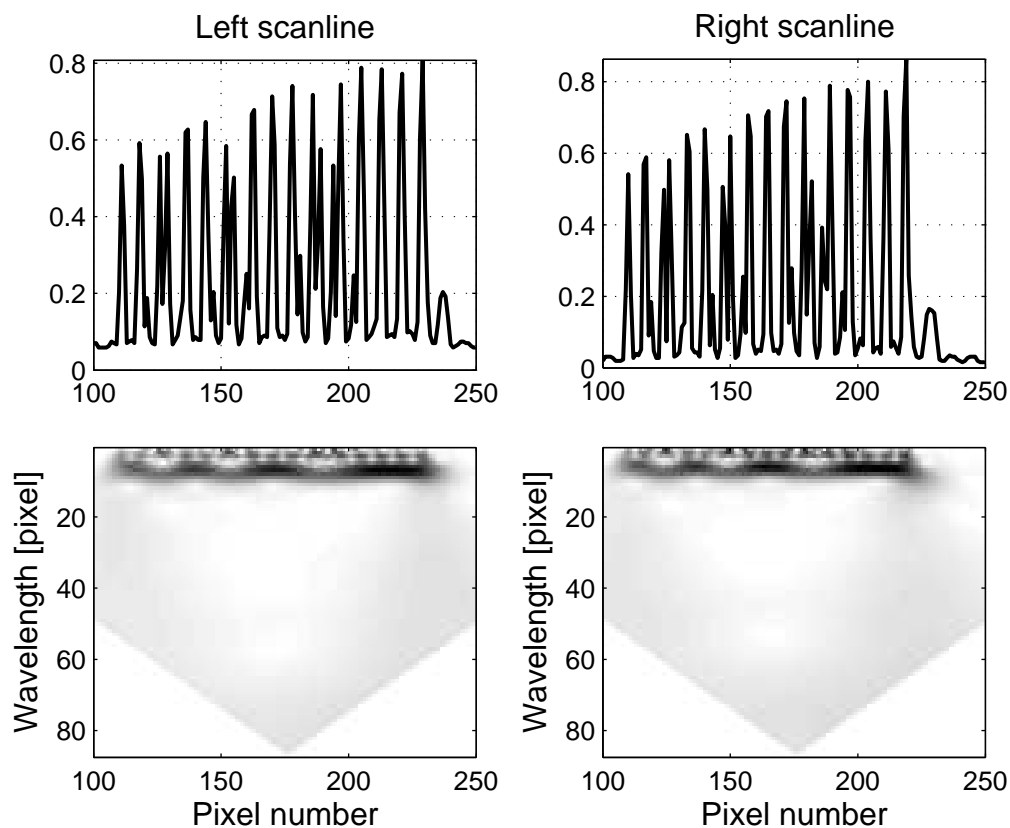
and then replace that in Eq.6.7, giving the final expression for the projected form (frequency shift) also known as sampling ratio:

$$\Psi = 1 + \frac{disp \tan \theta}{f - x_L \tan \theta} \quad (6.10)$$

This form Eq. 6.10 relates the parameters in the image plane to the surface slope θ and does require use of some known parameters (focal length f , image location x_L , and a candidate disparity $disp$).



(a) The effect of foreshortening on scalogram magnitude. Two views of flat surface with a sinusoidal texture appear on top, the scalogram magnitudes for their central scanlines appear below. The responses are similar, but are compressed to higher frequencies in the rotated view



(b) Left and right central scanlines of a flat surface tilted 45 degree in the top. Their corresponding scalogram shown in the bottom. We can see similar features in both scalograms; those in the right image are present at higher spatial frequencies because the right image is subjected to greater foreshortening than the right image

Figure 6.4: Foreshortening effects in the scalogram representation.

6.3 Extending Phase-based Algorithm

Phase-based methods such as [3] [34] [79] [2] as well as our method that is presented in chapter 5 can benefit from this foreshortening-factor to improve their performance.

The evaluation function Eq. 5.12 uses a global minimization strategy to find the best disparity from a list of candidates, so it is easy to incorporate a foreshortening correction term ψ to find the best matching and the disparity value. Moreover to searching disparity space, we also search over surface angle, which allows us to predict the corresponding frequency directly.

$$EF = \frac{1}{|\mathcal{U}|} \sum_{\lambda \in |\mathcal{U}|} \rho_l(c, \lambda) \cdot \left| \Delta_{\phi_{ideal}} - (\phi_l(c, \lambda) - \phi_r(c + \underline{disp}, \lambda \cdot \underline{\Psi})) \right|_{2\pi} \quad (6.11)$$

Pseudocode for the complete algorithm is given in Tab. 6.3, we can note the correction term (foreshortening-factor) Ψ on the right image phase measurements.

Table 6.1: Pseudo-code for the corrected Phase-based stereo algorithm

Given: A pair of grayscale images, list of candidate disparities, surface angles, and focal length f .

For each row
 Compute left and right scalogram S_L and S_R

For each column c
 $W_L(c) = \{\lambda: \rho_L(c, \lambda) \text{ exist and } \phi_L \text{ is reliable}\}$
 $W_R(c) = \{\lambda: \rho_R(c, \lambda) \text{ exist and } \phi_R \text{ is reliable}\}$

For each candidate disparity $disp$
For each angle θ
 $W = W_L(c) \cap W_R(c + disp)$
 $\Psi = 1 + \frac{disp \tan \theta}{f - x_L \tan \theta}$
 $EF = \frac{1}{|W|} \sum_{\lambda \in W} \rho(c, \lambda) \cdot \left| \Delta_{\phi_{ideal}} - (\phi_L(c, \lambda) - \phi_R(c + disp, \lambda \cdot \Psi)) \right|_{2\pi}$

Return disparity that yield minimum error, where denotes the foreshortening factor.

6.4 Results

The objective of this work is to demonstrate the effects of slanted surface. Therefore the correctness of our algorithm immediately becomes evident when dealing with the stereo pairs which is shown in Fig. 6.5. Left images of stereo pairs images "Flat

surface” rotated in 0° , 20° and 45° (depth increasing from right to left) are shown in Fig. 6.5 (a). Fig. 6.5 (b) show the reference disparity maps for each stereo pair images. These true maps obtained in our labs (at the university of Magdeburg), makes it possible to do measurements based on photogrammetric method, by applying a fast active 3-D measurement of geometrical shapes by photogrammetry and structured lighting method that was provided by [53]. In general a pattern of stripes caring intensity are projected onto the scenes to facilitate image point correspondence.

The computed disparity maps in Fig. 6.5 (c) shows that the disparities are linearly decreasing (*i.e. depth are linearly increasing*) from right to left (from bright to dark). Since the disparities are coded with brighter areas assigned as nearer, darker area assigned further away from the viewer. Therefore, the results showed how the proposed algorithm overcomes the problem of perspective foreshortening. The disparity map converts to 3-D surface based on the known camera geometry (focal length baseline), the visualization view of the 3-D surfaces reconstruction are shown in Fig. 6.5 (d). From the primary result from the visual comparison between the reference disparity maps and the computed disparity map we can observe the accurate and the smoothness of the computed results. A representative comparison of correspondence lines profile from reference disparity map (TrueDisp in dashed line) and the computed disparity map (CompDisp in solid line) is shown in Fig. 6.6. The right plots showed the amount of error in each pixel along the scan line.

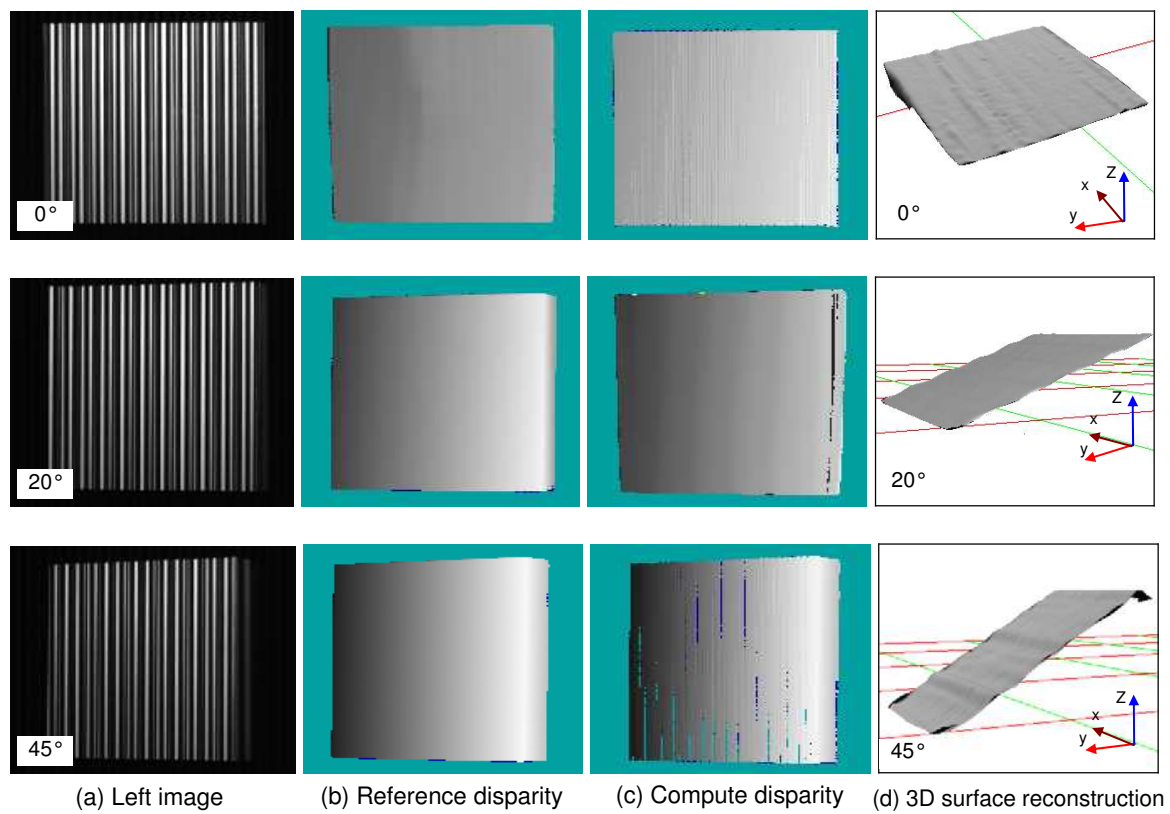
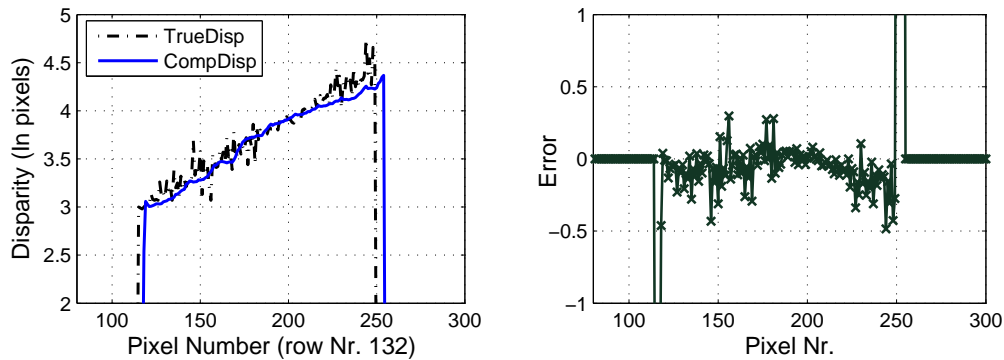
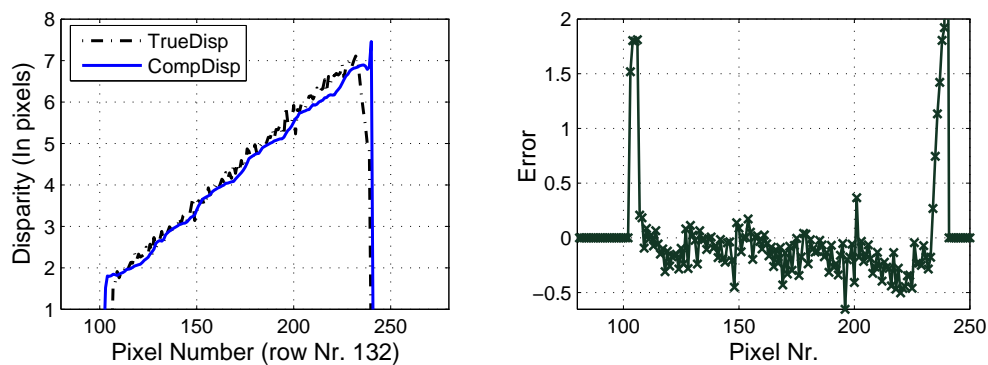


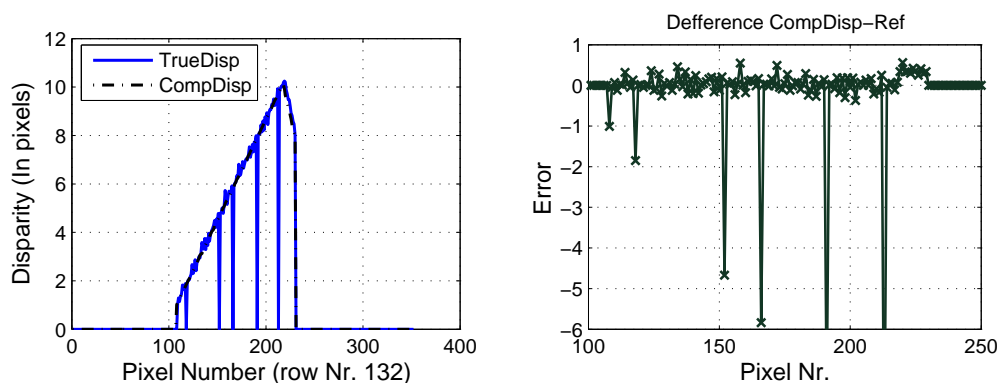
Figure 6.5: Results for slanted object. (a) are Flat surface rotated by angles 0° , 20° and 45° . (b) are the ground truth disparity maps (reference disparity). (c) are the 3-D surface reconstruction rotated for a good visualization.



(a) lines profile comparison from the computed (CompDisp) and the reference disparity (TrueDisp)



(b) lines profile comparison from the computed (CompDisp) and the reference disparity (TrueDisp)



(c) lines profile comparison from the computed (CompDisp) and the reference disparity (TrueDisp)

Figure 6.6: Ground truth compare against the computed disparity maps, for the center scanline of the flat surface at the various rotation angle; 0° , 20° and 45° and disparity range (1...6) pixels.

Chapter 7

Summary and Conclusions

This dissertation has provided a powerful analytical tool for image analysis in stereovision, that has addressed several problems; foreshortening, ambiguous matches, detecting and handling discontinuities and occlusion, and quantitative evaluation of stereo results.

Our contribution to the stereo problem has explored the benefits of local spatial frequency representation. There are two fundamental extremes in the study of 2-D imagery; the detail oriented *spatial view* and the wholistic *frequency view*. In the spatial view, an image is represented by the concatenation of independent pixel values. In the frequency view, the image is broken down mathematically into several frequency components, information in each component related to the image as a whole. Each view has its benefits; the spatial view can represent discontinuous textures and local segmentations directly, while the frequency view is a mathematically elegant representation that enables many useful analysis over large regions. In this research, a new approach called *local-spatial-frequency* approach is proposed to combine between the localizability of the spatial approach and the analytical benefits of the frequency approach.

While the stereovision is an attractive and a widely used method, it is rather limited to reconstruct the 3-D surface reconstruction, due to the correspondence problem. On the other hand, the correspondence problem can be reduced by using a method based on the structured light concept. Therefore, our experimental set-up consists of a combination of stereo vision concept and the structured light concept. Therefore our system contains a light source (projector) placed between two of cameras which are placed at different view points to capture the left and right images (stereo pair images) while the projector emits a light strips pattern into the scene. Thus increasing the local discriminatingly of each pixel and facilitating matching process. In order to simplify the matching process, rectifying the images is one of the important and initial steps in our method. By using the camera parameters that are obtained from the camera calibration process the captured stereo images are transferred to images as if obtained by camera with parallel optical axis. Therefore corresponding points

must always lie along epipolar lines in images. These lines correspond to the intersections of an epipolar plane (the plane through a point in scene and nodal points of the two cameras) with the left and right image planes. Exploiting this epipolar constraint reduces the initially 2-D search to a 1-D one.

Stereo matching is a technique used to extract depth from a pair of images that was taken from a slightly different view points. Many computation algorithms have been proposed [25], [2], [34], [41], [35], [36], [4], [26], [37]. These algorithms differ from one another in matching primitives, the density of the results, the accuracy of the estimates and the underlying computation time. In general, single step stereo algorithms can be classified in three classes: area-based, feature-based and frequency-based stereos. The difficulties with approaches based on area correlation are well known. Because of the difference in viewpoints, the effect of shading can give rise to differences in brightness for surfaces. A more serious difficulty arises from the effect of differing amount of foreshortening in the two views whenever a surface is not strictly fronto-parallel. Still another difficulty arises at surface boundaries, where a depth discontinuity may run through the region of the image being used for correlation. It is not even guaranteed in this case that the computed disparity will lie within the range of disparities present within the range.

This research has developed the frequency-base approaches, which has become a widespread method for depth and optical flow estimation. Because of its superior performance and better theoretical grounding. Therefore, we have provided some solutions of the long-standing problems in stereovision in two suggested algorithms:

The first algorithm [5], [6] is based on using the output of linear spatial filters, which is presented for characterizing the information present in a vector of filter responses. The left and right images convolving with a set of linear filters tuned to a number of different orientations and scales (using a Gaussian function and its derivatives as basis functions). The responses of these filters at a given point constitute a vector that characterizes the local structure of the image patch. The correspondence problem can be solved by seeking points in the other view where this vector is maximally similar.

Our contribution in this algorithm is to develop this filter-based framework, by presenting a technique that exploit the constraints arising from viewing geometry. A general viewing geometry is assumed, with the optical axis converged at a fixation point, instead of the simpler case of parallel optical axes frequently assumed in machine vision. This technique is described by some processing levels, whereby the first level deals with filter design and optimization. This contains the Gaussian derivatives, filter design, evaluate and optimize filter degree using Steering theorem and singular value decomposition (SVD). The second processing level of the technique is specified by correspondences analysis, which describes the feature extraction; filter response vector (FRV) and estimating the disparity map. The improvement of the disparity takes place by the use of an adaptive scale filter, where a substantial area of

the filter is applied to a region of significant depth variation, such as inappropriately large scale filter should selectively ignored.

The method provides a rich description of the image, with little chance of false matching. The foremost contributions arise in detecting and handling the depth discontinuities and the occlusion region, where unpaired points lie in regions seen only in one view. This algorithm maintains a current best estimate of the viewing parameters (to constrain vertical disparity to be consistent with epipolar geometry), a visibility map (to record whether a point is binocularly visible or occluded) and a scale map (to record the largest scale of filter not straddling a depth discontinuity). The computational complexity of this algorithm has two significant terms. The first one is the cost of the initial linear spatial filtering at multiple scale and orientations. Implementations can be made quite effect by using separable kernels and pyramid strategies. The second term is corresponding to the cost of computing the disparity map. This cost is proportional to the number of iterations.

The second suggested algorithm [7], [8] is a phase-difference based algorithm that provide a solution for long-standing problems; perspective foreshortening and ambiguous matches in stereo vision, without needed to do any iteration. We also demonstrate by example the utility of the local spatial frequency representation in the context of stereo vision.

The algorithm combined the magnitude and phase information for estimating depth information from two-dimensional stereoscopic image pairs. This method takes into account not only the instability of phase but also the surface perspective distortion (the foreshortening in one view). These properties are important to the use of phase information in order to avoid the incorrect disparity estimates. Instead of matching intensities directly, a Gabor scale-space expansion is used. Magnitude information is used to detect "weak points" in the frequency domain, and only reliable phase values remain for a robust estimation disparity. This method provides a foreshortening correction factor to overcome the perspective distortion region, and demonstrates a novel solution to a phase-wraparound problem that has limited the application of other phase-based method. The advantage of this algorithm is that the disparities values estimates are obtained with sub-pixel accuracy without requiring explicit sub-pixel signal reconstruction. It also relates the parameters in the image plane to the surface slope and does not require prior knowledge of the distance to the object or an estimate of the disparity derivative. The experimental results show that the performance of the proposed algorithm in terms of accuracy and density of the disparity estimates has greatly improved. The foreshortening factor has its greatest impact when objects are sharply slanted and located near the cameras. The computed results are evaluated with the ground truth.

Some particular contributions include:

Perspective Foreshortening

We addressed the long-standing problem of perspective foreshortening in stereo vision. Our method provides an analytical closed-form expression for the effect of perspective foreshortening on stereo matching in the frequency domain. It also demonstrate that the model improves the results of a real stereo method. Results relating foreshortening to the frequency domain have appeared in the shape-from-texture literature [66], but without a description of the use of disparity, the mapping between the two areas was left as a general affine matrix. Similarly, [41] uses an affine matrix to relate two corresponding image patches in the context of stereo, but does not provide the analytic relationship. We provided a presentation to unify these three terms (stereo, foreshortening, and frequency domain) in both the world model and the stereo images.

Adaptive Scale Selection

We demonstrated the improvement of our scale-adaptive algorithm over traditional multiscale (e.g., coarse to fine) algorithms. We also presented a multiscale phase-based method that is not confused by missing information at intermediate scales. Most of the prior stereo systems work either at a single scale or using a coarse-to-fine approach in a fixed order. Our system invoke automatic scale selection (via filter magnitude weighting) in a nonrestrictive manner and can handle missing information at any scale. Sanger [34] has used a magnitude weighting, but imposed a restrictive limit on candidate disparities. Jones and Malik [41] also used some scale-space processing, but only to eliminate the coarsest scales at depth discontinuities.

Phase Wraparound

We eliminated the restriction of previous phase-based stereo algorithms on the maximum disparity range, and described a new stereo algorithm that eliminates the problem of phase wraparound.

Data reference

We provided a fast active 3-D measurement of geometrical shapes by photogrammetry and structured lighting method. That was presented in our group [53], [54] which obtain a reference data set. Wherever, at our Lab in the university of Magdeburg, it is possible to make measurements based on photogrammetric method. In this method [53], [54], instead of correlating area in two images, greyscale values of single pixel obtained from a number of different images taken in succession are correlated. Thus, it is possible to achieve higher spatial resolution.

The computed disparity results are compared with Scharstein and Szeliski data set [35] to obtain a comparison with various well-known stereo matching algorithms. That are provided in the Web-site "cat.middlebury.edu/stereo" [1].

In order to obtain the algorithm ranking, the four different stereo pair images that found in [1] with their ground truth are used. The statistic based on the percentage of bad pixels. Therefore, to compute these statistic over whole image the evaluation focus on three different kinds of regions: all pixels in non-occluded regions ($B_{\bar{o}}$), all pixels in half-occluded regions (B_{ho}) and all pixels near-occluded regions i.e. near discontinuities (B_{no}). As the taxonomy and evaluation methodology which presented in [35]. Our phase-difference based algorithm is compared with 31 others existing algorithms. The percentage of bad pixels, (the pixels which deviate is more than 0.75 unit from the true disparity are labeled as "*bad pixels*"), each algorithm is sorted according to its overall ranks. For all errors our algorithm stands in the upper half of the table.

The computed disparity maps also are compared against the ground truth by computing the percentage of wrong pixels. A pixel is judged to be erroneous, if its absolute deviation from the ground truth is larger than 0.75 unit. The statistics results for all data sets in a quantitative measures to evaluate the quality of the computed results: The percentage of bad matching pixels, Root mean square error, average error for the nonsingular points, the mean percentage relative error, the deviation of the average error and the percentage number of correct pixels are reported. In addition,

A representative comparison of correspondence lines profile from reference disparity map and the computed disparity map show the accurate and the smoothness of the computed result. The random error could be determined by measuring reference (ground truth), for instance in the experimental results of cylinder object the error amounts to approximately ± 0.2 pixels. Also for a center scanline from the slanted flat surface at the various rotation angles; 0° , 20° and 45° and disparity range (1...6) pixels, The error amounts are approximately ± 2.5 pixels.

Potential future extensions to this work include:

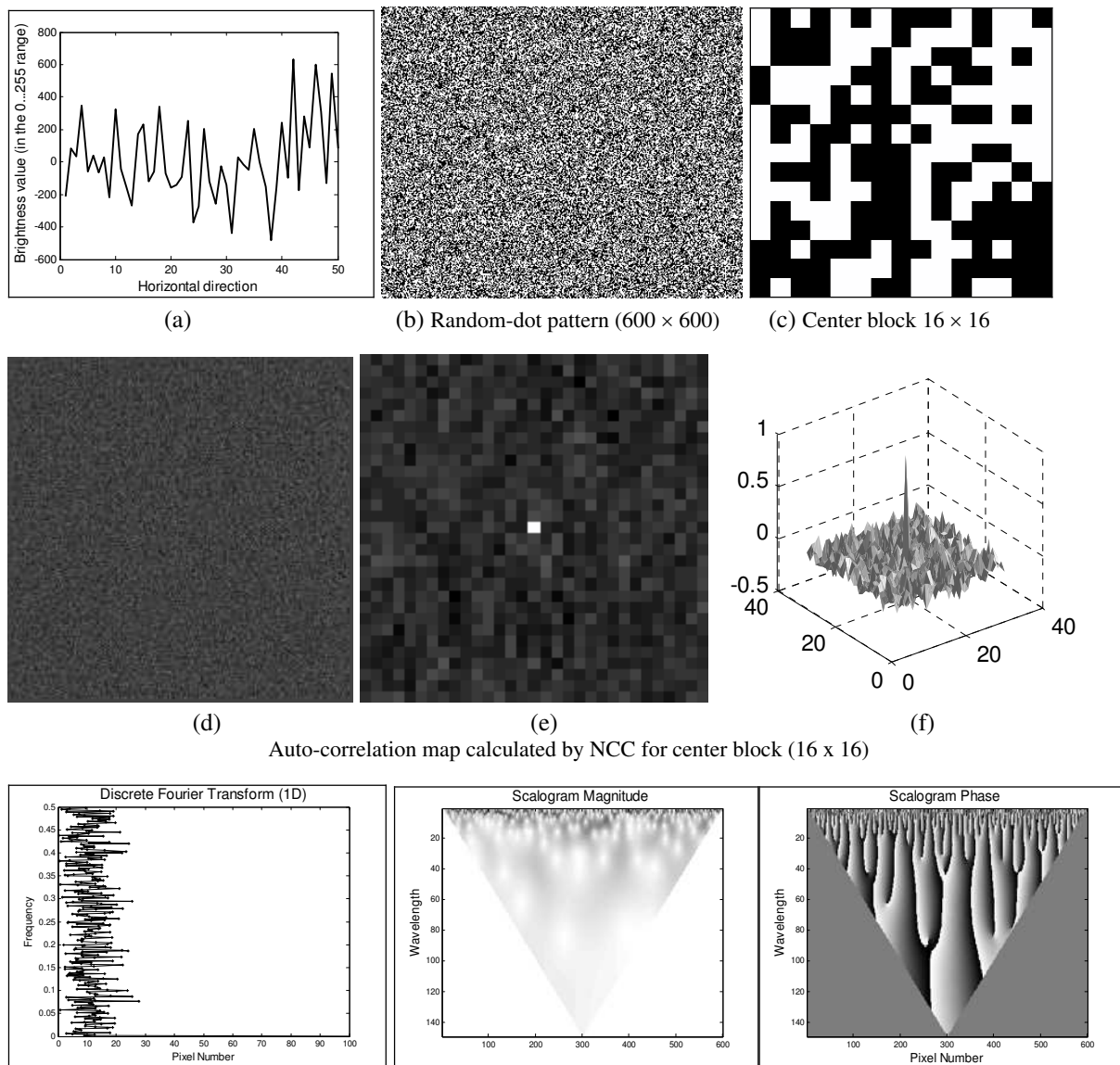
Experimentation with the shape of the evaluation function (as in Equation 5.12); a replacement for AbsDiffMod such as a cosine might yield smoother results, or be faster to implement. The other point related to speed up the processing by incorporating the fast wavelet transform with appropriate interpolation in place of the complete scalogram computation, using a smaller set of filters, and a smaller set of foreshortening angle candidates.

Appendices

Appendix A

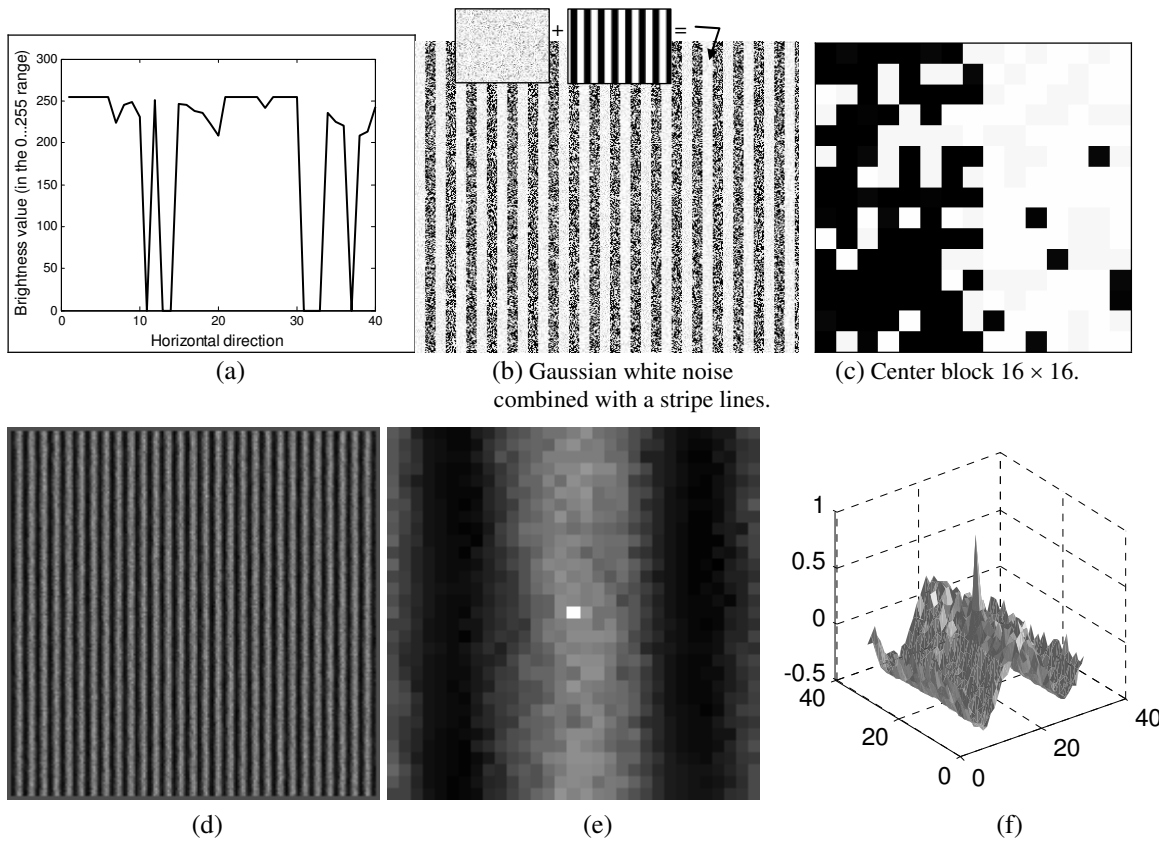
Structured Light Pattern

In this appendix, we review the structured light patterns that were analyzed in chapter 2; Random-dot pattern, combined Gaussian white noise with light stripes pattern with default parameters (zero mean, 0.01 variance), Checker-board pattern, Sinusoidal pattern with a period length 80 pixels, Sawtooth pattern with a period 80 pixels, and Color stripes pattern generated as a sequence of light stripes. In each figure; graph (a) shows a square wave to illustrate the concept of spatial frequency resolution where spatial frequency refers to how rapidly the brightness signal is changing in space, graph (b) shows the originally 600×600 structured lighting pattern, a central block 16×16 depicted in graph (c), a distribution map of auto-correlation values calculated by NCC function using a 16×16 pixel region in the center part of the pattern depicted in graph (d), (e) and (f). The goal of using such structured light patterns is to increase the ability of local discriminant at each pixel and the facility of the matching process.

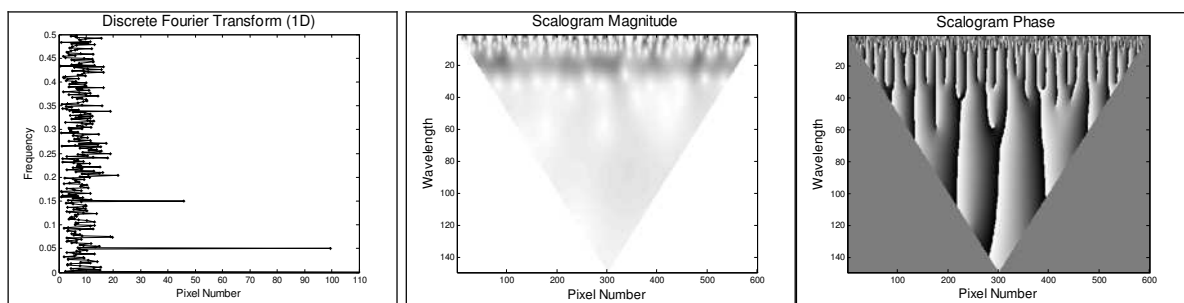


Spatial-frequency response for a one line profile from the pattern. The discrete Fourier transform and the scalogram (Magnitude and Phase) representation calculated as Eq. 3.5

Figure A.1: Random-dot pattern 600×600 , the brightness value in the range 0...255.

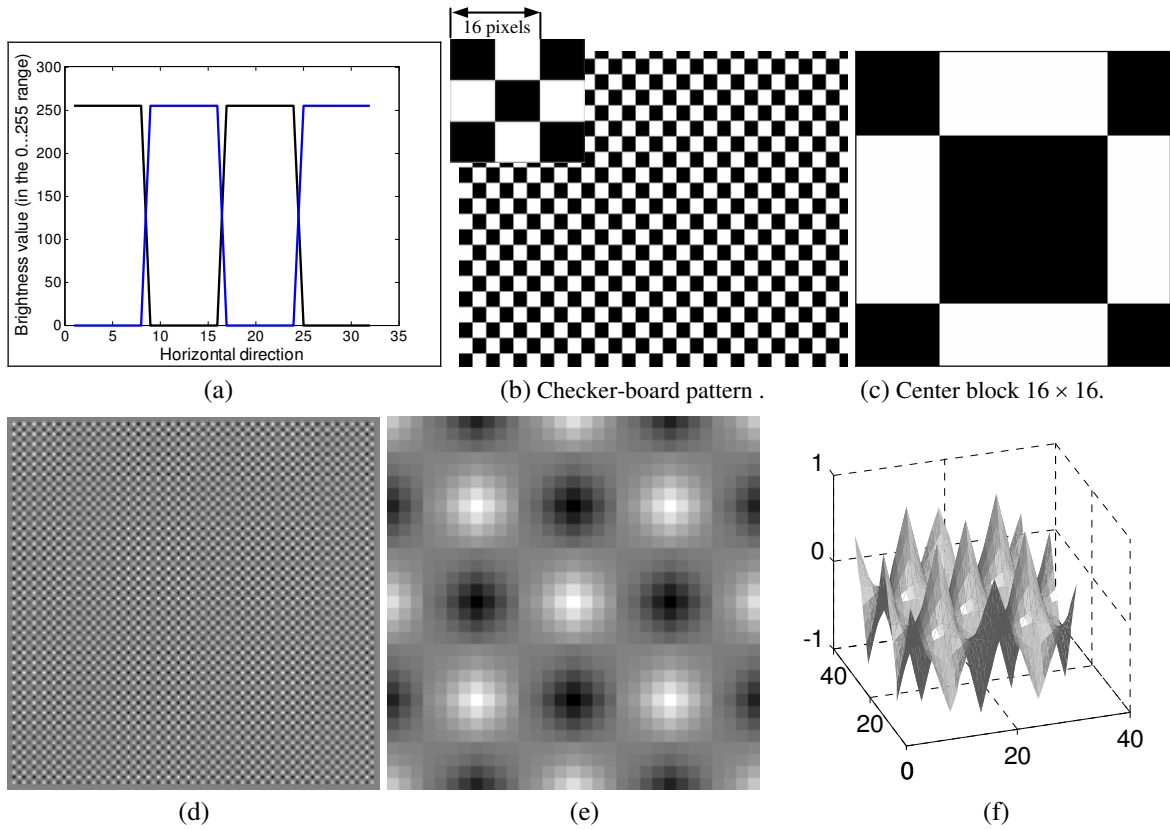


Auto-correlation map calculated by NCC for center block (16×16)

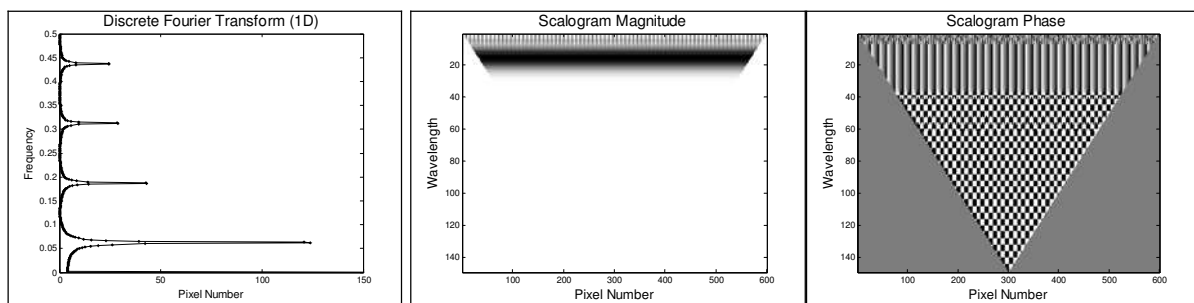


Spatial-frequency response for a one line profile from the pattern. The discrete Fourier transform and the scalogram (Magnitude and Phase) representation as description in chapter 3, Eq. 3.5

Figure A.2: A combined Gaussian white noise in default parameters; zero mean noise and 0.01 variance with a stripe line of width 8 pixels.

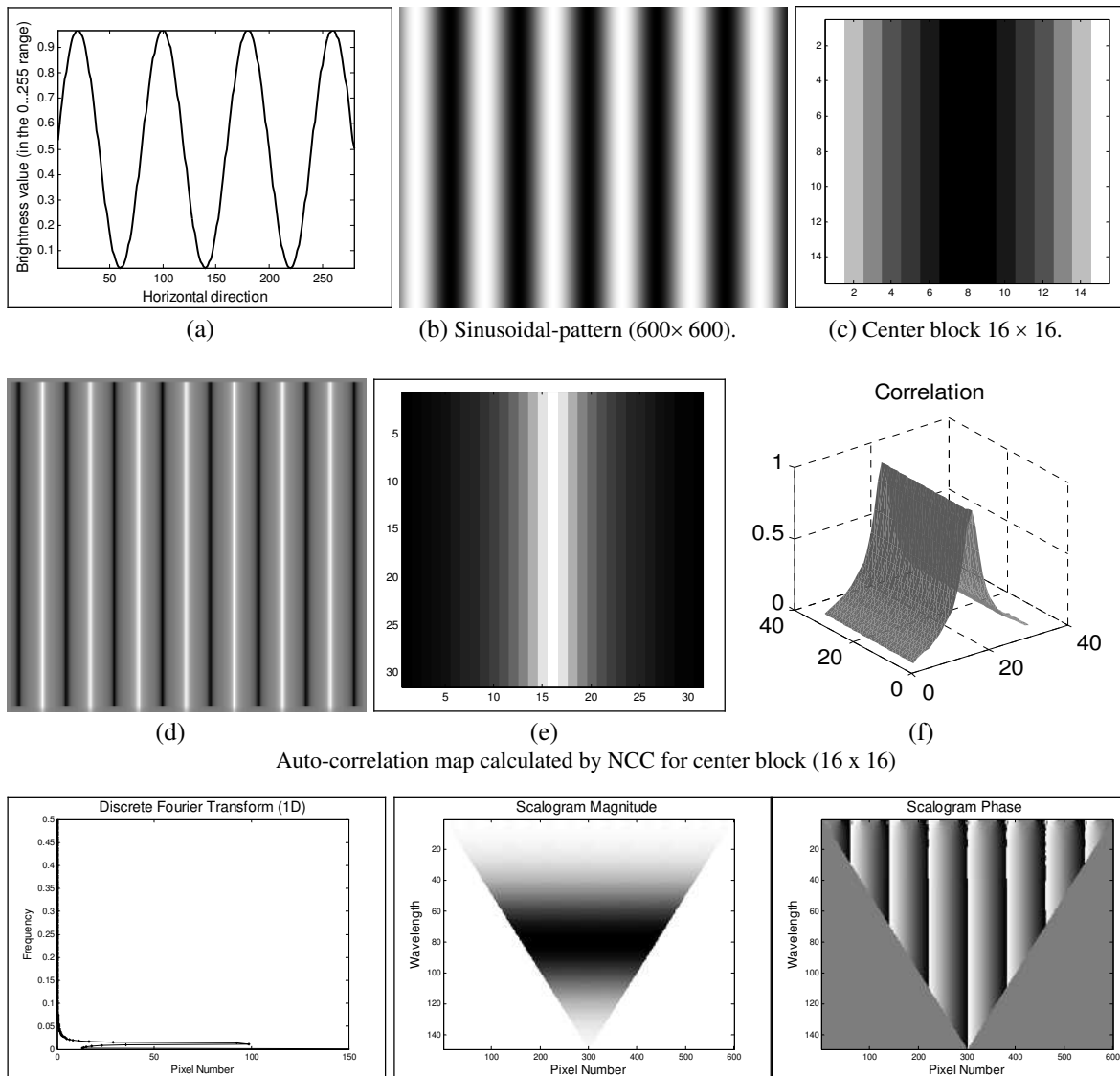


Auto-correlation map calculated by NCC for center block (16 x 16)



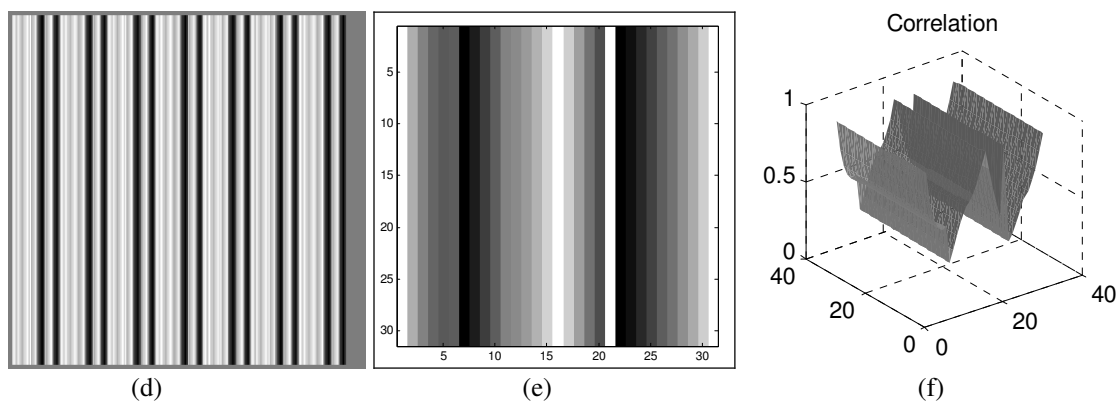
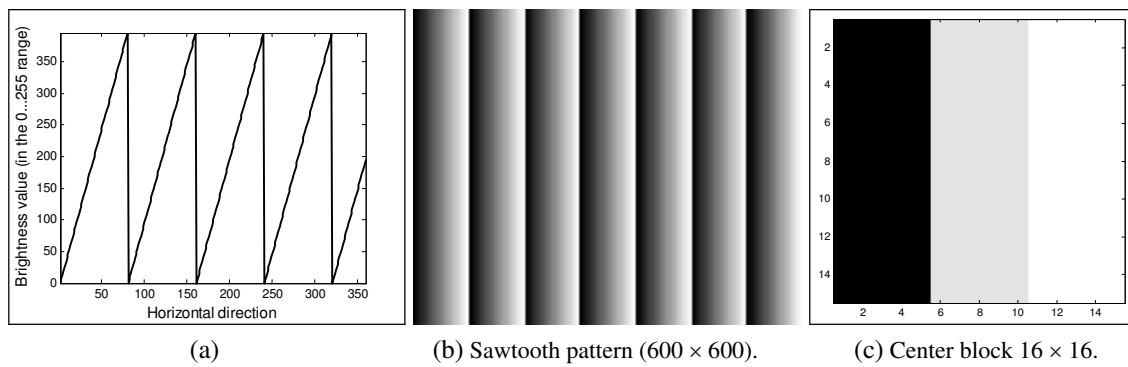
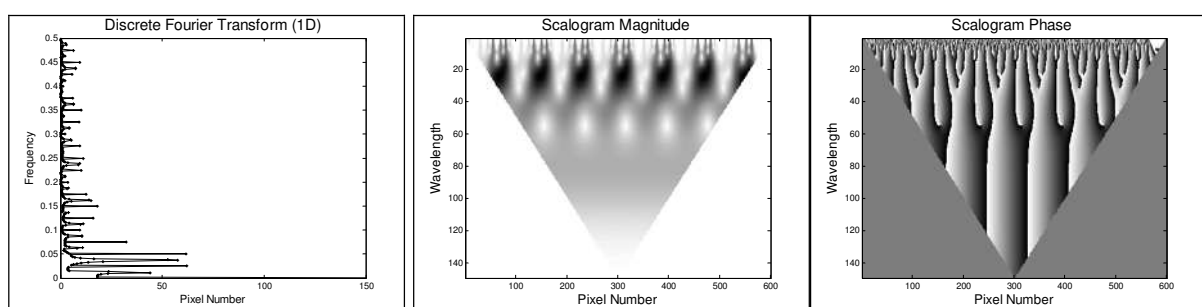
Spatial-frequency response for a one line profile from the pattern. The discrete Fourier transform and the scalogram (Magnitude and Phase) representation as description in chapter 3, Eq. 3.5

Figure A.3: Checker-board pattern, with a basic structure alternating bright and dark squares.



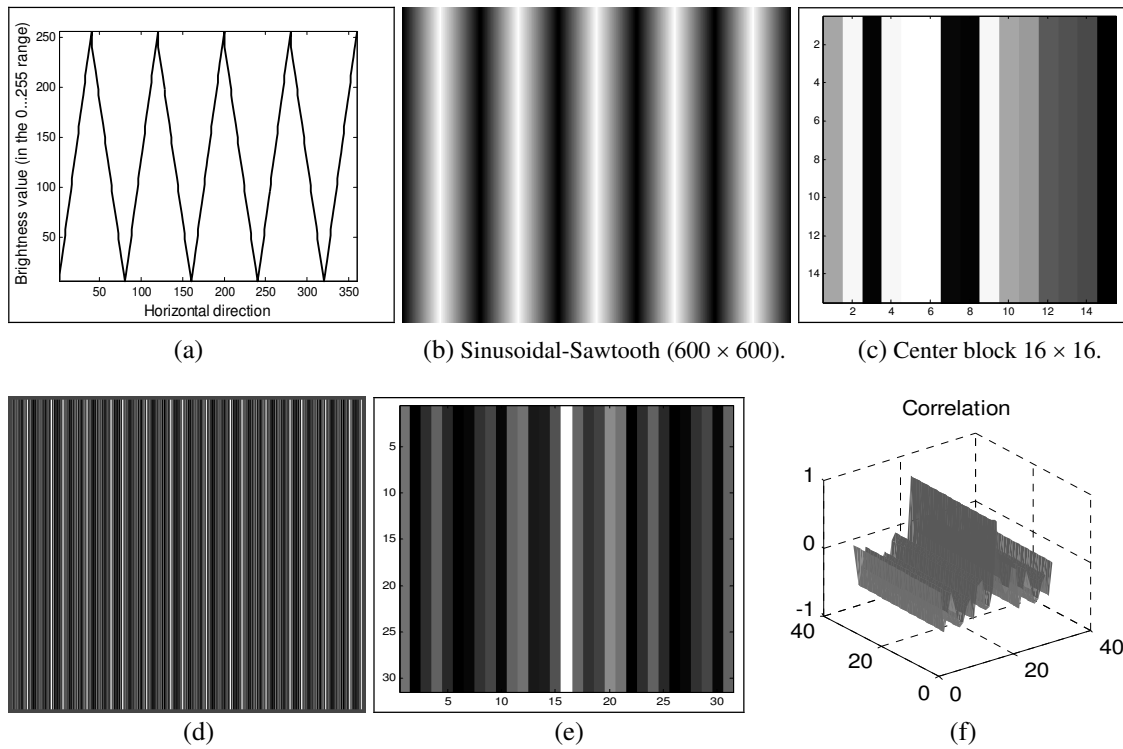
Spatial-frequency response for a one line profile from the pattern. The discrete Fourier transform and the scalogram (Magnitude and Phase) representation as description in chapter 3, Eq. 3.5

Figure A.4: Sinusoidal pattern with a period length 80 pixels.

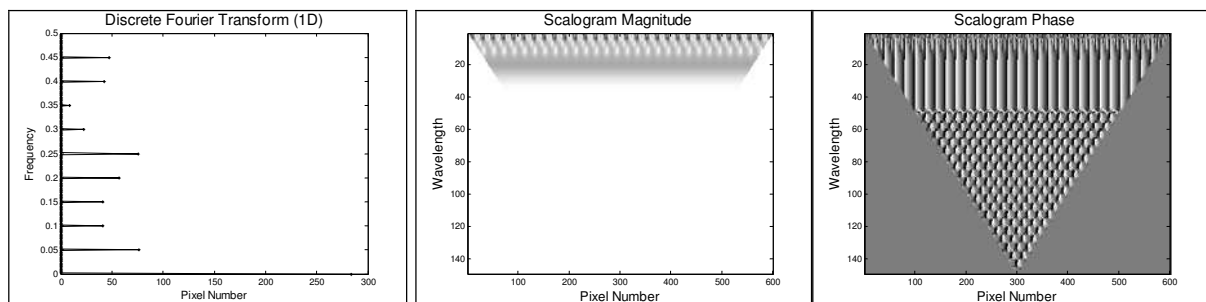
Auto-correlation map calculated by NCC for center block (16×16)

Spatial-frequency response for a one line profile from the pattern. The discrete Fourier transform and the scalogram (Magnitude and Phase) representation as description in chapter 3, Eq. 3.5

Figure A.5: Sawtooth pattern with a period 80 pixels

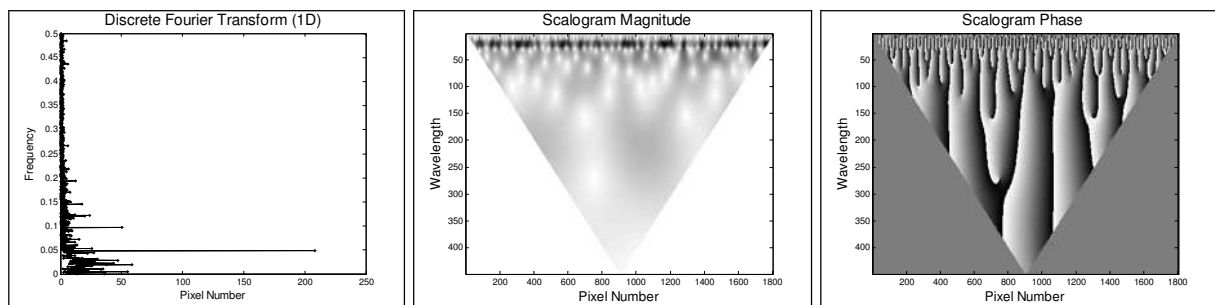
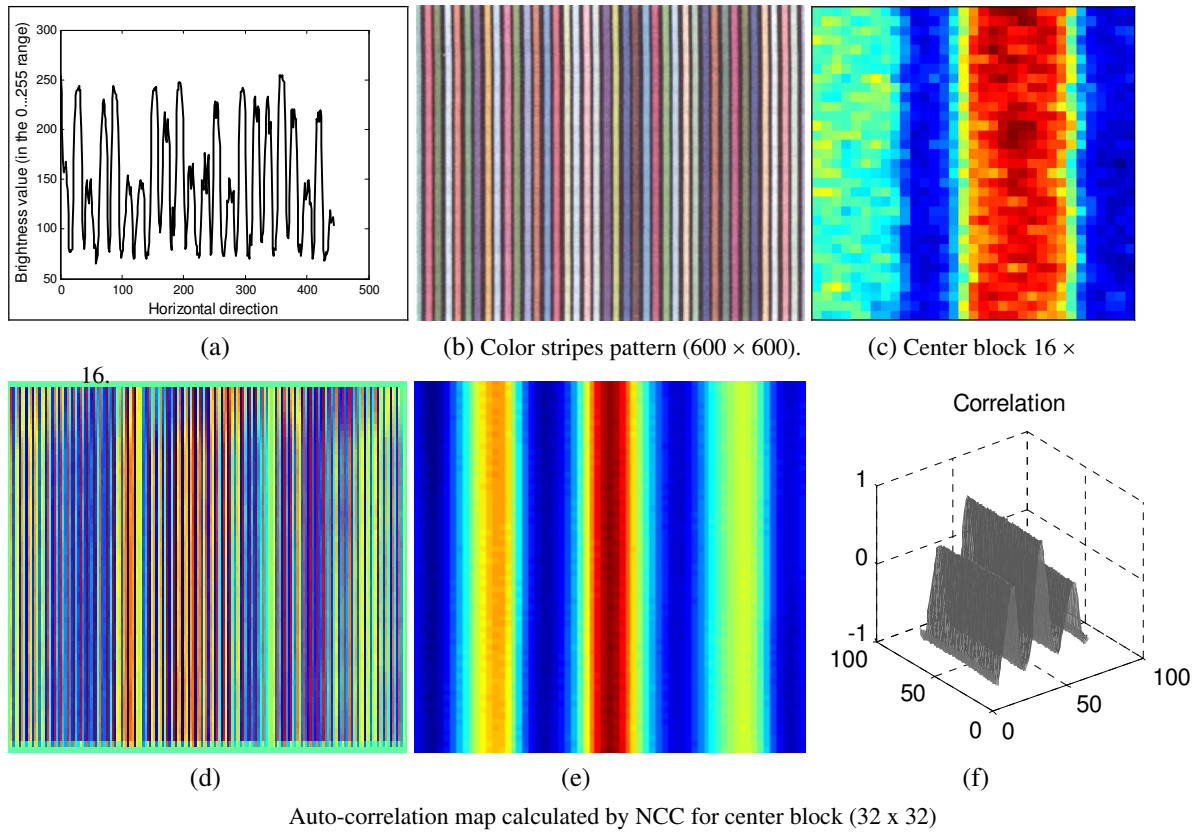


Auto-correlation map calculated by NCC for center block (16 x 16)



Spatial-frequency response for a one line profile from the pattern. The discrete Fourier transform and the scalogram (Magnitude and Phase) representation as description in chapter 3, Eq. 3.5

Figure A.6: Sinusoidal-Sawtooth pattern with a period 80 pixels



Spatial-frequency response for a one line profile from the pattern. The discrete Fourier transform and the scalogram (Magnitude and Phase) representation as description in chapter 3, Eq. 3.5

Figure A.7: Color stripes pattern generated as a sequence of light stripes with brighter colors, $C_1, B, C_2, B, \dots, B, C_N$, a black stripe B is inserted between each C_i and C_{i+1} to increase the intensity variation between adjacent light stripes.

Bibliography

- [1] D. Scharstein and R. Szeliski, "<http://vision.middlebury.edu/stereo/>."
- [2] U. Ahlvers, U. Ahlvers, and U. Zlzer, "Inclusion of Magnitude Information for Improved Phase-based Disparity Estimation in Stereoscopic Image Pairs," *Proc. Int. Conf. on Computer Vision*, pp. 3–9, 2005.
- [3] D. J. Fleet, A. Jepson, and M. Jenkin, "Phase-based disparity measurement," *CVGIP*, 53(2), pp. 198–210, 1991.
- [4] J. Weng, "Image matching using the windowed Fourier phase," *International Journal of Computer Vision archive*, Vol. 11 No. 3, pp. 211–236, 1993.
- [5] S. El-Etiby, A. K. Al-Hamadi, and B. Michaelis, "Improvement the Accuracy of 3-D Reconstruction Considering Distortion in Stereovision using a Set of Linear Spatial Filters," *IEEE International Conference on Computational Intelligence for Modelling Control and Automation-CIMCA '2005*, pp. 654–660, Vienna, Austria 2005.
- [6] S. El-Etiby, A. Al-Hamadi, and B. Michaelis, "Improvement of 3-D Reconstruction by Deformable Stereo Matching using a Set of Linear Spatial Filters," *International Journal on Graphics, Vision and Image Processing, 2006, Special Issue on Image Retrieval and Representation, ISSN [1687-398X]*, pp. 41– 46, 2006.
- [7] S. El-Etiby, A. K. Al-Hamadi, and B. Michaelis, "Dense Depth Map Reconstruction by Phase Difference-based Algorithm Under Influence of Perspective Distortion," *Special Issue of the International Journal Machine Graphics and Vision MGV 2006*, 2006.
- [8] S. El-Etiby, A. Al-Hamadi, and B. Michaelis, "Dense stereo correspondence with slanted surface using phase-based algorithm."
- [9] B. Triggs, P. F. Mclauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle Adjustment - A Modern Synthesis," *ICCV '99: Proceedings of the International Workshop on Vision Algorithms*, pp. 298–372, 2000.

-
- [10] J. Heikkila and O. Silve, "Calibration procedure for short focal length off-the-shelf CCD cameras," *Proc. 13th Intl. Conf. on Pattern Recognition*, pp. 166–170, 1996.
- [11] J. Heikkila and O. Silven, "A Four-step Camera Calibration Procedure with Implicit Image Correction," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'97)*, pp. 1106–1112, 1997.
- [12] R. K. M. Pollefeys and L. J. V. Gool, "A simple and efficient rectification method for general motion," *In proceeding international conference on computer vision, ICCV 1999*, pp. 496–501, 1999.
- [13] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications*, 12(1), pp. 16–22, 2000.
- [14] A. A. Goshtasby, *2-D and 3-D Image Registration for Medical, Remote Sensing, and Industrial Applications*, 1st ed. Wiley-Interscience; 1 edition, ISBN-10: 0471649546, March 8, 2005.
- [15] R. Klette, K. Schluens, and A. Koschan, *Computer Vision: Three-Dimensional Data from Images*, 1st ed. Springer, 1998.
- [16] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v.16 no.9, pp. 920–932, September 1994.
- [17] D. Geiger, B. Ladendorf, and A. YUille, "Occlusion and binocular stereo," *ECCC*, pp. 425–433, 1992.
- [18] A. Fusiello, V. Rberto, and E. Trucco, "Efficient stereo with multiple windowing," *CVPR*, pp. 858–863, Juni, 1997.
- [19] A. F. Bobick and S. Intille, "Large occlusion region," *IJCV*, Vol. 33 No. 3, pp. 181–200, Sept., 1999.
- [20] H. Tao, H. Sawhney, and R. Kumar, "A global matching framework for stereo computation," *ICCV,1*, pp. 532–539, 2001.
- [21] J. Mulligan and K. Daniilidis, "Predicting disparity windows for real-time stereo," *Lecture notes in computer science,1842*, pp. 220–235, 2000.
- [22] Y. Ohta and T. Kanade, "Stereo by intra-and inter-scanline search using dynamic programming," *IEEE PAMI7*, pp. 139–154, 1984.
- [23] M. Irani and P. Anandan, "About direct methods," *In Vision Algorithms: Theory and Practice. Springer-Verlag*, 2000.

-
- [24] P. H. S. Torr and A. Zisserman, "Feature Based Methods for Structure and Motion Estimation," *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms, Springer Berlin-Heidelberg, Vol. 1883/2000*, pp. 278–294, 2000.
- [25] U. Ahlvers, U. Zlzer, and S. Rechmeier, "FFT based Disparity Estimation for Stereo Image Coding," *Proc. IEEE (ICIP'03)*, vol. 1, pp. 761–764, 2003.
- [26] E. Lilienblum and B. Michaelis, "Optical 3D Surface Reconstruction by a Multi-Period Phase Shift Method," *Journal of Computers (JCP)*, vol. 2, pp. 73–83, December April 2007.
- [27] F. Zhao, Q. Huang, and W. Gao, "Image Matching by Normalized Cross-Correlation," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006, Vol. 2*, May 2006.
- [28] T. Baumbach, "Shift Detection by Restoration Signalbasiertes Bildmatching und spezielle Anwendungen," 2000.
- [29] R. Calow, "Markerlose Ganganalyse mit einem Multikamerasystem," Ph.D. dissertation, Otto-von-Guericke-University Magdeburg, IESK, Germany, 2005.
- [30] D. J. Fleet and A. Jepson, "Stability of phase information," *IEEE Transaction on pattern analysis and machine, intelligence*, Vol. 15 No. 12, pp. 1253–1268, 1993.
- [31] A. Cazzi and B. Crespi, "Performance of phase-based algorithms for disparity estimation," *International journal of Machine Vision and Applications Vol. 9, Issue 5-6*, pp. 334–340, March-1997.
- [32] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, 194, pp. 283–287, 1976.
- [33] G. Medioni and R. Nevatia, "Segment-based stereo matching," *Computer Vision, Graphics and Image Processing*, 31, p. 218, 1985.
- [34] T. D. Sanger, "Stereo Disparity computation Using Gabor Filter," *biological Cybernetics*, pp. 405–418, 1988.
- [35] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms (SSD + min-filter)," *International Journal of Computer Vision*, 47, pp. 7–42, 2002.
- [36] C. Strecha, R. Fransens, and L. V. Gool, "Combined depth and outlier estimation in multi-view stereo," *CVPR 2006*, pp. 2394–2401, 2006.

-
- [37] R. Niese, A. Al-Hamadi, and B. Michaelis, "A Novel Method for 3D Face Detection and Normalization," *JOURNAL OF MULTIMEDIA*, ISSN: [1796-2048, 2007], 2007, VOL. 2, NO. 5, SEP. 07, pp. 1–12, 2007.
- [38] J. Kim and J. Park, "New stereo matching and 3D view generation algorithms using aerial stereo images," pp. 663–669, 2004.
- [39] L. Falkenhagen, "Hierarchical block-based disparity estimation considering neighbourhood constraints," *International workshop on SNHC and 3D Imaging*, pp. 115–122, Rhodes, Greece, 1997.
- [40] C. Kuglin and D. Hines, "The phase correlation image alignment method," *Proceedings of the IEEE Int. conf. on cyber. Society*, pp. 163–165, 1975.
- [41] D. Jones and J. Malik, "Determining three-dimensional shape from orientation and spatial frequency disparities," *European Conference on Computer Vision, ECCV 1992*, pp. 662–669, 1992.
- [42] P. M. Will and K. S. Pennington, "Grid Coding: A preprocessing technique for robot and machine vision," *In proceeding international journal conference on artificial intelligence*, pp. 66–70, 1971.
- [43] Y. Shirai and M. Suwa, "Recognition of polyhedrons with a range finder," *In proceeding international joint conference on Artificial intelligence*, pp. 80–87, 1971.
- [44] G. J. Abin and T. Binford, "Computer Description of Curved Object," *In Proceeding International Joint Conference on Artificial Intelligence*, pp. 629–640, 1973.
- [45] H. Yamamoto, K. Sato, and S. Inokuchi, "Range imaging System Based on Binary Image Accumulation," *In Proceeding International Conference on Pattern Recognition*, pp. 233–235, 1986.
- [46] Y. Sato, H. Kitagawa, and H. Fujita, "Shape measurement of Curved Objects Using Multiple Slit-Ray Projections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 4, No. 6, pp. 641–646, 1982.
- [47] K. Kemmotsu and T. Kande, "Uncertainty in Object Pose Determination with Three Light-Stripe Range Measurements," *IEEE International Journal of Robotics and Automation*, Vol. 11, No. 5, pp. 741–747, 1995.
- [48] M. Asada, H. Ichikawa, and S. Tsuji, "Determining surface orientation by projecting a stripe pattern," *Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume 10, Issue 5*, pp. 749 – 754, Sep 1988.

- [49] Y. F. Wang, "Characterizing Three-Dimensional Surface Structure from Visual Images," *PAMI*, Vol. 13, No. 1, pp. 52–60, January 1991.
- [50] C. S. Chen, Y. P. Hung, C. C. Chiang, and J. L. Wu, "Range Data Acquisition Using Color Structured Lighting and Stereo Vision," *Image and Vision Computing*, Vol. 15, pp. 445–456, June 1997.
- [51] S. B. Kang, J. A. Webb, and C. L. Zitnick, "A Multibaseline Stereo System with Active Illumination and Real-time Image Acquisition," *IEEE Fifth International Conference on Computer Vision (ICCV 1995)*, pp. 88–93, June 1995.
- [52] M. Hashimoto and K. Sumi, "3-D Object Recognition Based on Integration of Range Image and Gray-scale Image," *BMVC 2001*, 2001.
- [53] P. Albrecht and B. Michaelis, "Improvement of the Spatial Resolution of an Optical 3-D Measurement Procedure," *IEEE Transactions on Instrumentation and Measurement*, Vol. 47 No. 1, pp. 158–162, 1998.
- [54] ———, "Stereo Photogrammetry with Improved Spatial Resolution," *Proceedings of the 14th International Conference on Pattern Recognition*, Vol. I, pp. 845–849, 1998.
- [55] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 891–906, 1991.
- [56] C. Chun-Ming and L. Andrew, "Coherence of multiscale features for enhancement of digital mammograms," *IEEE Trans. on Infor. Technology in Biomedicine* 3(1), pp. 32–46, 1999.
- [57] D. Kalman, "A Singular value decomposition: the SVD of a Matrix," *College Mathematical Journal*, Vol. 27, No. 1, pp. 2–23, 1996.
- [58] R. A. Young, "The GAussian derivative model for spatial vision: I. Retinal mechanisms," *Spatial Vision*, Vol. 2, pp. 273–293, 1987.
- [59] J. A. Bloom and T. R. Reed, "A Gaussian Derivative-Based Transform," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 551–553, 1996.
- [60] W. E. Michael, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," in *A Practical Approach to Microarray Data Analysis*. D.P. Berrar, W. Dubitzky, M. Granzow, eds., LANL LA-UR-02-4001, pp. 91–109, 2003.
- [61] L. Jacobson and H. Wechsler, "Joint spatial/spatial-frequency representation," *Signal Processing*, vol. 14, no. 1, p. 3768, 1988.

-
- [62] A. Beghdadi and R. Iordache, "Image quality assessment using the joint spatial/spatial-frequency representation," *EURASIP J. Appl. Signal Processing* 2006, Article ID 80537, 2006.
- [63] R. Olivier and Martin, "Wavelets and signal processing," *IEEE Signal processing Magazine*, pp. 14–38, October (1991).
- [64] R. R. abiner and R. W. Schafer, *Digital Processing of Speech Signals*, prentice-hall ed. Englewood Cliffs, NJ 1978.
- [65] J. Krumm and S. Shafer, "Local spatial frequency analysis of image texture," *Proceedings of the Third International Conference on Computer Vision*, pp. 354 – 358, December 1990.
- [66] J. Krumm, "Space Frequency Shape Inference and Segmentation of Textured 3D Surfaces," *Robotics Institute, Carnegie Mellon University*, December 1993.
- [67] D. Gabor, "Theory of Communication," *Journal IEE*, 93, pp. 429–459, 1946.
- [68] J. Tribolet, "A new phase unwrapping algorithm," *IEEE Transaction on Acoustics, Speech, and Signal Processing*, Vol. 25 No. 2, pp. 170–177, 1977.
- [69] B. Crespi and G. Tecchioli, "Adaptive Gabor Filters for Phase-Based Disparity Estimation," *International journal of pattern recognition and artificial intelligence* Vol. 13, No. 5, pp. 591–614, 1999.
- [70] Anonymous, "Three ways to improve stereo subpixel accuracy for low-textured regions," *Submitted to ICCV 2007*, 2007.
- [71] H. Hirschmueller, "Stereo vision in structured environments by consistent semi-global matching," *CVPR-2006*, pp. 2386–2393, 2006.
- [72] V. Kolmogorov and R. Zabih, "Computing Visual Correspondence with Occlusions using Graph Cuts," *IEEE International Conference on Computer Vision (ICCV-2001)*, pp. 508–515, July (2001).
- [73] Anonymous, "A progressive edge-based correspondence method," *ICCV 2007*, 2007.
- [74] Q. Yang, L. Wang, R. Yang, H. Stewnius, and D. Nistr, "Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling," *CVPR-2006*, pp. 2347–2354, 2006.
- [75] K. J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *PAMI*, 28(4), pp. 650–656, 2006.

- [76] M. Gong and Y. H. Yang, “Near real-time reliable stereo matching using programmable graphics hardware,” *CVPR, Vol I*, pp. 924–931, 2005.
- [77] Anonymous, “Stereo vision enabling precise border localization within a scanline optimization framework,” *Submitted to ACCV 2007*, 2007.
- [78] H. Audirac, A. Beloiarov, F. Nez, and J. Villegas, “Dense disparity map based on STICA algorithm,” *Expo Forestal*, 2005.
- [79] J. Weng, “A theory of image matching,” *Third international conference of computer vision, IEEE computer society*, pp. 200–209, 1990.

Curriculum Vitae

Name:	Sherif S. A. El-Etriby
Date of Birth:	August 10, 1970; in Tanta, Egypt
Nationality:	Egyptian
Status:	Married, three Children
Address:	Universitätsplatz 1, D-39106 Magdeburg, Germany
E-mail:	Sherif.El-Etriby@E-Technik.Uni-Magdeburg.de

Education:

1986 - 1989	Secondary School Certificate (Higher secondary certificate), Tanta, Egypt.
1989 - 1993	B.Sc. in Pure mathematics and computer science. From Minufiya University, Egypt.
1995 - 1996	Postgraduate courses for M.Sc. Faculty of science, Minufiya University, Egypt.
1997 - 1999	M.Sc. in Mathematics and computer science. From Minufiya University, Egypt.
2003 - Now	Works towards Ph.D. degree at the Institute of Electronics, Signal Processing and Communications (IESK), Otto-von-Guericke University Magdeburg, Germany.

Work Experience:

1994 - 1999	Demonstrator in Dept. of Mathematics and Computer Science, Faculty of Science, Minufiya University, Egypt.
1999 - 2002	Assistant lecturer in Dept. of Mathematics and Computer Science, Faculty of Science, Minufiya University, Egypt.
2002 - 2003	Assistant lecturer in Dept. of Computer Science, Faculty of Computers and Information, Minufiya University, Egypt.
2003 - Now	PhD student, Chair of Technical Computer Science, Otto-von-Guericke University, Magdeburg, Germany.

Magdeburg, 7.January.2008

Sherif El-Etriby

Related Publications

The presented thesis is based on the following international reviewed journal and conferences papers:

1. El-Etriby S., Al-Hamadi A. K., Michaelis B.: **”Compositional character of 3-D surface reconstruction using phase-difference technique”**, Journal of Machine Vision and Applications, ISSN: 0932-8092 (Print version), ISSN: 1432-1769(electronic version), Springer.(Submitted)
2. Elmezain M.; Al-Hamadi A.; El-Etriby S.; Michaelis B. : **Gesture Recognition for Alphabets from Hand Motion Trajectory Using Hidden Markov Models**. The 7th IEEE International Symposium on Signal Processing and Information Technology, ISSPIT-2007 , pp. 1209-1214, 15-18,December 2007.
3. El-Etriby S.; Al-Hamadi A.; Michaelis B. :**Dense stereo correspondence with slanted surface using phase-based algorithm**. IEEE International Symposium on Industrial Electronics, Vigo, Spain, IEEE ISIE, June 2007. pp. 1807-1813, CD, ISBN: [1-4244-0755-9].
4. El-Etriby S.; Al-Hamadi A.; Michaelis B. :**Dense Depth Map Reconstruction by Phase Difference-based Algorithm Under Influence of Perspective Distortion**. Special Issue of the International Journal ”Machine Graphics and Vision MGV-2006 vol. 15, no. 3/4, pp. 349-361”, ICCVG 2006, ISSN: [1230-0535].
5. El-Etriby S.; Al-Hamadi A.; Michaelis B. : **Phase-Based Disparity Estimation with Implicit Foreshortening-Correction**. The 4th International Multi conference on Computer Science and Information Technology CSIT 2006, IEEE, Amman -Jordan, pp.510-520.
6. El-Etriby S.; Al-Hamadi A.; Michaelis B. : **Improvement of 3-D Reconstruction by Deformable Stereo Matching using a Set of Linear Spatial Filters**. International Journal on Graphics, Vision and Image Processing, 2006, Special Issue on Image Retrieval and Representation, ISSN [1687-398X], pp. 41-46.

7. El-Etriby S.; Al-Hamadi A.; Michaelis B. : **Improvement the Accuracy of 3-D Reconstruction Considering Distortion in Stereovision using a Set of Linear Spatial Filters.** IEEE International Conference on Computational Intelligence for Modelling Control and Automation - CIMCA'2005, Vienna, Austria, pp.654-660.
8. El-Etriby S.; Al-Hamadi A.; Michaelis B. : **Improvement of 3-D Reconstruction by Deformable Stereo Matching using a Set of Linear Spatial Filters.** ICGST International Congress for global Science and Technology, Conference on Graphics, Vision and Image Processing (GVIP-05), Cairo 2005, pp. 445-450