3D Range Data Acquisition Using Structured Lighting and Accuracy Phase-Based Stereo Algorithm

Sherif El-etriby

Faculty of Computers and Information, Menoufia University-MUFIC, Egypt
sherif.el-etriby@ci.menofia.edu.eg

Abstract

The model of local spatial frequency provides a powerful analytical for image analysis. In stereo matching algorithms, we develop phase difference-based algorithm that use an adaptive scale selection process at the corresponding points in the two views. This takes into account surface perspective distortion (foreshortening problem). Challenges arise from the fact that stereo images are acquired from a slightly different view. Therefore, a surface image is more compressed and occupied a smaller area in one view. Instead of matching intensities directly, a Gabor scale-space expansion (scalogram) is used. The phase difference at corresponding points in the two images is used to estimate the disparity. The central ingredient of the proposed model is the development of the dual scale factor that allow us to reason about foreshortening effects in both geometric domain of the world model and frequency domain of the stereo images. The method is based on the convolution of the stereo image pair with Gabor-scale space expansion and caching the outputs in a local spatial frequency, that known as Scalogram representation. The proposed method provides a constraint to detect and filtering out the singularity points with its neighborhoods. Moreover, it provides a novel solution to the phase wraparound problem that has limited the application of other phase-based methods. The advantage of this algorithm is that the computed disparity values are obtained with sub-pixel accuracy without requiring explicit sub-pixel signal reconstruction. The suggested algorithm provides an analytical closed-form expression for the effect of perspective foreshortening. Wide-ranging evaluation methodologies used to confirm the efficiency and performance on the basis of analysis of rectified real images. Hence, the proposed method has a superior performance in comparison with other methods.

Keywords: Stereo Matching, Structured Lighting, Local Spatial Frequency.

I. INTRODUCTION

Stereo matching is an algorithm used to extract depth from a pair of images that was taken from slightly different viewpoints. 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.

The task is to match primitives of the input stereo images, thereby solving the correspondence problem, and then determine the depth (the 3rd-Dimention) of objects in the scene. It remains a difficult problem to solve because of several reasons; the stereo has inherent ambiguities, image noise, depth discontinuities and occlusion region, as well as the complexity of 3-D scene which is often viewed from a geometric perspective distortion, one of the most significant problem know as Unequal Projection Lengths (Foreshortening problem); the object surfaces are acquired in a slanted view, therefore, they appear more compressed in one image, due to perspective foreshortening. Figure 1 illustrates this problem where a projection line segment \( N \) from left camera and a projection line segment \( M \) from right camera have a different number of pixels, this effect makes matching its two images very difficult, since its appearance differs so much between the two images.

In this paper we attempt to overcome some of the long-standing problems in stereovision by using structured lighting and the frequency-based (phase-different based) algorithm. That 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), 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. Subsequently, a novel technique is introduced that reduces phase nonlinearity by means of an adaptive mechanism for the tuning frequency.
In contrast with other stereo algorithms, the proposed lighting-stereo algorithm can produce artificial textures on the object surfaces by projecting a stripe pattern. Hence, the stereo matching problem can be solved in a more stable way. The structured pattern is designed to contain many sharp edge, these sharp edges are utilized to improve the accuracy of the disparity map as well as the 3D surface reconstruction.

The experimental set-up for the 3D measurements (IESK, Magdeburge Uni, German) is shown in Fig.2. 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. And an ABW LCD-640 (Automatisierung Bildverarbeitung) stripe projector with a maximum resolution of 640×640 is used to generate arbitrary line patterns.

![Figure 2. Experimental Setup: The Two PULNix Dual-Tap AccuPiXEL cameras (stereo camera) ABW LCD-640 Active light projector.](image)

The 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, [12] proposed a slit line projection to recognize curvilinear object. In 1986 Yamamoto [13] proposed a half plane illumination system instead of a slit line. Over the last few years, a large number of techniques have been developed [14], [2], [8], [5], [15], [1], [17] etc.

Kemmotsu [18], use a method based on the projection of two or three slit lines with different orientation and position in the 3-D coordinates system. Chen [19] has 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 [20] proposed a 4-camera system to recover dense stereo range data from each set of images, a sinusoidal varying pattern project onto the scene to enhance local intensity discriminability.

Scharstein and Szeliski [15] 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. [6] 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. J. Kim [9] presented a new stereo matching algorithm based on window warping technique in hierarchical matching process to balance the perspective distortions. L. Falkenhagen [10] 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.

In the frequency-based algorithm, not only phase but also magnitude information is available to have 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 [11], which utilizes the phase shift theorem of the Fourier transform. The method propose a model based on the inverse Fourier transform of the phase difference between two images. Sanger [8] 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.

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, Weng [4] used the windowed Fourier phase, which is the multiple-window using the Kuglin-Hines algorithm [11]. The window size is allowed to vary so that the measurements can be more localized. Ahlvers [14], [2] presented a new approach for combining magnitude and phase information in Fast Fourier Transform (FFT)-based algorithms.

II. STRUCTURED LIGHTING AND RANGE DATA ACQUISITION

The projection of light patterns into a scene is called "Structured lighting". The light source can be a spatial light modulator or a video projector. 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 correspondence problem can be reduced using a method based on the structured light concept, then the distance of an object to the camera (the 3-D coordinates) can be determined through analyzing the observed light patterns in the images. Figure 3 shows the principle of structured lighting and stereovision system. The system contains a projector (light source) and two cameras, a light pattern projected into the scene. The origin of the stereo camera coordinates system is \(X'Y'Z'\) and \(XYZ\). The considered illuminated object point \(P(X'Y'Z')\) is formatted to a point \(P(x, y, z)\) in the left image plane as well as point \(P(x, y)\) in the right image plane.

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 viewpoints to capture the left and right images. Thus increasing the local discriminability of each pixel and facilitating matching process. 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.

A. Structured Pattern Generation

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

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.

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. Therefore, \(2^n\) light planes are uniquely encodable by using \(n\) patterns, i.e. taking \(n\) images. Then a large number of images have to be generated for the object.

B. 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.

From figure 3 slightly location \((x_i, y_i)\) of a pixel \(p_i\) in the image constrains the 3-D location of the corresponding object point \(P(X_i, Y_i, Z_i)\) 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 \((X, Y, Z)\) is obtained.

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. Obviously, 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.

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 over-come 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. 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.

The effect of varying the frequency (period length) of the stripe pattern is depicted in Fig 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, 0.4cm, and 0.2cm) respectively in Fig. 5(a), the 3D point cloud of the computed disparity map in each case shown in Fig 5(b) and the 3-D surface reconstruction shown Fig 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).

III. LOCAL SPATIAL FREQUENCY REPRESENTATION

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 view (spatial and frequency view) and provides information of localized space-frequency events.

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 measures the frequency of the sinusoid signal.

When filtering a periodic signal there are two common techniques for determining the frequency of the signal: first, using the filter frequency directly is complicated because any discrete filter will have an ambiguous response and this assumption may reduce the precision of the results. Second, the derivative (instantaneous frequency) used to overcome some complementary problems in most phase based stereo methods.

In proposed method, the stereo images are transformed from the space to the frequency domain. Therefore, the disparity is shown as a phase difference between the image signals. Many problems in stereovision arise from its limited image representation. There are many local frequency representations; spectrograms (Short Time Fourier Transforms (STFT)), Wigner-Ville distributions and the adaptive filters (scalogram).

The scalogram is one of the local frequency representations, which uses variable window size that shrinks and grows as the tuning frequency changes along the scanline on the image. This makes high frequencies much more localizable, and provides the necessary support for low frequencies (e.g., using more low frequency filters). The scalogram is actually a special case of the wavelet with a Gabor function as the transfer function. It illustrates a relatively higher-frequency component exists at the ends of the signal, while a low-frequency part occurs in the middle. This localization is the power of scalogram representation, since the sampling is linear in wavelength.

For generating of the scalogram, \( S_d(x, \omega) \) 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 \( \rho \) and phase \( \phi \), that compute as follows:

Let \( G(x, \omega) \) be the function defining a Gabor filter centered at the origin, and \( R(x) \) is the current image row.

The scalogram representation i.e. the spatial convolution between \( G(x, \omega) \) and the signal (scan line) \( R(x) \) is calculated as:

\[
S_d(x, \omega) = R(x)^* G(x, \omega) \tag{1}
\]

\[
S_d(x, \omega) = \int_{-\infty}^{\infty} R(\tau)G(x - \tau) d\tau \tag{2}
\]

Here the symbol * denotes the convolution operation and the Gaussian envelope of \( G(x, \omega) \) define the local neighborhood. \( R \) is a one-dimensional row in the input image; \( x \) is a central point of the kernel window.

Figure 6 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 6(c) 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.

IV. PHASE-DIFFERENCE AS DISPARITY

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 multiscale 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.

Let the left and right image \( F_l(x) \) and \( F_r(x) \) be sinusoids wave, each of them have a particular phase angle and the frequency is \( \omega_l, \phi_l \) and \( \omega_r, \phi_r \) 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 equivalent. Then the disparity \( \Delta x \) at any point in the signal is a different distance at that position \( x_l - x_r \) given by:
\[ \Delta x = x_f - x_r = \frac{\phi_r - \phi_f}{2\pi \lambda} \]  

(3)

This is called the constant frequency model. In particular, the phase measurement translates directly into disparity measurement. But, problems arise when applying this to the discrete case. However, how actual frequency is measured? And 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 contributions is manipulate this situation by providing a foreshortening correction factor based on the physical geometry of the scene.

V. EXACTNESS COMPLEMENTARY PROBLEMS

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. We propose an automatic technique to locally adapt the filter parameters to the input signal, and overcome the most common problems in phase-based stereo algorithms: Singularity neighborhoods, Phase warparound, and precision effects of the Foreshortening).

Figure 7 shows the concept of the suggested algorithm. It will be described in several levels; the primary step is the preprocess the image rectification. Therefore corresponding points must always lie along epipolar lines in images. These lines correspond to the intersections of an epipolar with the left and right image planes. 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. Finally, the 3-D coordinates of the examined object points are obtained using conventional triangulation method.

A. Singularity Neighborhoods

An advantage of phase information is its stability with respect to contrast differences between the left and right views. In some regions, the phase signal has an unwanted behavior, thus scalogram is analytic, and contains a number of isolated zeros, where \( S_p(x,\omega) = 0 \), that appears as white spots in the magnitude representation. The phase signal is undefined when the magnitude fades away hence disparity will not be computed accurately, these points known as phase singularities.

Those are marked by the arrows in scalogram-magnitude Fig. 8(b). 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.

Now, we describe the characteristic behavior of \( \rho(x, y) \) and \( \phi(x, y) \) near singular points. Let \( P(x_0, \lambda_0) \) denotes the location of a singularity where the singularity point lies near the center at the point which the phase contours intersect, see Fig. 9. 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 \( (2\pi/\lambda) \). Within these neighborhoods regions exist in which local frequency is zero; i.e., \( \frac{\partial \phi(x,y)}{\partial x} = 0 \). 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.

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

\[
\sigma(\lambda) \left| \frac{\partial \rho(x, \lambda_0)}{\partial x} + \frac{1}{\rho(x, \lambda_0)} \frac{\partial \phi(x, \lambda_0)}{\partial x} - \frac{2\pi}{\lambda} \right| < \tau
\]  

(4)

As \( \tau \) 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".

This constrain ignores the unreliable phase values and uses only the remaining left and right phases to fit the disparity, (set cutoff for the threshold (\( \tau \)) that is used here is 0.05).

![Scalogram representation. The scalogram for a given row \( R \) is represented as a two dimensional matrix of complex numbers phase \( \phi \) and magnitude \( \rho \) are shown in subfigures A) and B) respectively.](image-url)
between the phase-difference and the disparity, \( \text{disp} \). We enumerate the given list of the possible disparity range specified by its tuning frequency \([3] [21]\). For more details, only be able to estimate disparity less than the wavelength (i.e. \( \text{disp} \)). Therefore, disparities of 1, 21, 41,... are appear equivalent, \( \text{disp} \). That means a given filter will only be able to estimate disparity less than the wavelength specified by its tuning frequency).

For more accuracy, we also should manipulate of phase-wraparound, wherever we can only measure the phase difference module \( 2\pi \) (i.e. a given filter will only be able to estimate disparity less than the wavelength specified by its tuning frequency). In this section we describe a process that eliminates the problem of phase wraparound. The difficulty lies in the fact that phase can only be measured modules \( 2\pi \). The problem of phase wraparound. The difficulty lies in the fact that phase can only be measured modules \( 2\pi \). Wherever we can only measure the phase value is expected to number of times that phase value is expected to number of times that phase can only be measured modules \( 2\pi \). Wherever we can only measure the phase difference module \( 2\pi \).

Therefore, (6) can be written as:

\[
\Delta \phi_{\text{ideal}} = \frac{2\pi \cdot \text{disp}}{\lambda}
\]

\[
\Delta \phi_{\text{meas}} = \phi_i(c, \lambda) - \phi_f(c + \text{disp}, \lambda)
\]

(5)

(6)

The measured phase-difference may be not equal the ideal-phase difference, since a wavelength, \( \lambda = 20 \). Therefore, disparities of 1, 21, 41,... are appear equivalent, as not expected. i.e. \( \Delta x(1) = \Delta x((\lambda + 1)), i = 1; 2; ..., \text{etc.} \)

\[
\Delta \phi_{\text{meas}} = |\Delta \phi_{\text{ideal}}| 2\pi
\]

Subsequently, for correcting the phase angle radian, one must add multiples of \( \pm 2\pi \), then \( \Delta \phi_{\text{ideal}} = \Delta \phi_{\text{meas}} + k2\pi \), which mean that the measured phase difference is only part of the results. In order to compute 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 \( k = 0 \). Other direct phase methods [7] [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. 11 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, (6) can be written as:

\[
\Delta \phi_{\text{meas}}(c, \lambda, \text{disp}) = \phi_i(c, \lambda) - \phi_f(c + \text{disp}, \lambda)
\]

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.

\[
EF = \frac{1}{|\lambda|} \sum_{k=0}^{\min(|\lambda|, 2\pi)} |\Delta \phi_{\text{meas}}(c, \lambda, \text{disp}) - \Delta \phi_{\text{ideal}}(c, \lambda, \text{disp})|
\]

(9)

The \( \lambda \) is the set of wavelengths whose output are considered reliable, and the absolute difference module (ADM) is the smallest difference between the ideal-phase and the measured phase, such that;

\[
\text{ADM}(a, b) = \min_{k \in [-1, 0.5]} ||a + k2\pi - b||
\]

(10)

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 (\( \Delta \phi_{\text{meas}} \)) and (\( \Delta \phi_{\text{ideal}} \)). Then we select the estimating disparity that exhibits minimum error as the result for the pixel.

C. Precision Effects of Foreshortening

This section describes the effect of perspective foreshortening in terms of local spatial frequency. To simplify the analysis, we assume the only object in the world plane is a textured at flat surface that is either parallel to the image plane, or rotated about the vertical axis by angle \( \theta \). We further assume that the stereo cameras have parallel optical (depth) and vertical (height) axes. We restrict our attention to the effects of foreshortening in one-dimensional image Scanlines. Figure 12 shows an overhead schematic 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.
The frequency shift between images of an oriented surface at surface is constant, it is independent of the surface texture, and can be expressed using only disparity and surface angle. Therefore, it must first determine how the appearance of the object’s surface texture changes between the perspective image planes. Therefore, it is useful to know at first how the sampling rate varies between the two images. For each distance $X_0$ along the world surface, we want to compare the projection segments in the left and the right images, $X_L$ and $X_R$, respectively, i.e., compare the left sampling rate ($\delta X_L / \delta X_L$) to the right sampling rate ($\delta X_R / \delta X_L$), consequently;

$$\text{sampling ratio} = \frac{\delta X_L}{\delta X_L} + \frac{\delta X_S}{\delta X_L} = \frac{\delta X_R}{\delta X_L} \quad (11)$$

This sampling ratio will call the foreshortening factor, which is denoted as $\psi$. As a result, this formula can compute the sampling ratio in the image space, without having to explicitly model the distance $X_0$ 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 since the purpose of the work is to estimate disparities values, it would be better if the disparity and its derivative could be avoided in the calculation of the foreshortening factor.

From the foreshortening model Fig. 12(left), the point varies across the surface, which can lead to the changes of its projection on the image plane. The relation between the projection points in the image plane (index pixel) and the surface angle is calculated using similar triangles.

$$\text{index pixel} = \frac{X_0 \text{world coordinate}}{Z_0 \text{world coordinate}}$$

Therefore, we obtain expressions for $X_L$ and $X_R$:

$$\frac{X_L}{f} = \frac{Z_L + X_S \sin \theta}{Z_L \cos \theta}, \quad \frac{X_R}{f} = \frac{Z_L - B}{Z_L \cos \theta} + \frac{X_S \cos \theta - B}{Z_L \sin \theta} \quad (12)$$

As a result, the equivalences expression for $X_S$ from the left and right camera geometry written as:

$$X_S = \frac{x_L Z_L}{f \cos \theta - x_L \sin \theta}, \quad \text{or} \quad X_S = \frac{x_R Z_L + B_f}{f \cos \theta - x_R \sin \theta} \quad (13)$$

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

$$\frac{x_L Z_L}{f \cos \theta - x_L \sin \theta} = \frac{x_R Z_L + B_f}{f \cos \theta - x_R \sin \theta} \quad (14)$$

Solving (14) for the right pixel index $x_R$ gives:

$$x_R = x_L \left[ 1 - B_f Z_L \right] - \frac{B_f Z_L}{Z_L} \quad (15)$$

Let us return to the foreshortening factor (11) and reduce the derivative by substituting for $x_R$.

$$\psi = \frac{\delta X_L}{\delta X_L} \left[ 1 + \frac{B_f}{Z_L} \right] - \frac{B_f}{Z_L} \quad (16)$$

This formula gives the geometric form of the foreshortening factor. To eliminate the distance in front of the left camera $Z_L$, the well-known relationship $\text{disp} = x_L - x_R$, can be rewritten as:

$$\text{disp} = \frac{B_f}{Z_L} - \frac{x_L \tan \theta}{Z_L} \Rightarrow \frac{B_f}{Z_L} = \frac{\text{disp}}{f - x_L \tan \theta} \quad (17)$$

and then replace that in (16), giving the final expression for the projected form:

$$\psi = 1 + \frac{\text{disp} \tan \theta}{f - x_L \tan \theta} \quad (18)$$

This form (18) relates parameters in image plane to the surface slope $\theta$, and dose require use of some known parameters (focal length $f$, image location $x_L$, and a candidate disparity $\text{disp}$). The evaluation function (9) uses a global minimization strategy to find the best disparity from a list of candidates, so it is easy to incorporate a foreshortening correction ($\psi$) in (9) to find the best matching as shown in (19).

$$EF = \frac{1}{|A|} \sum_{\lambda \epsilon A} \rho(c, \lambda) \cdot \left[ \Delta \phi_{\text{ideal}} - (\phi_c(c, \lambda) - \phi_\lambda(c + \text{disp}, \lambda \cdot \psi)) \right]_{2\pi} \quad (19)$$

VI. 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 and comparison with other algorithms.

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

| TABLE I. ATTRIBUTES OF FOUR STEREO PAIRS |
|-----------------|-----------------|-----------------|-----------------|
| Image pair      | Size            | Disparity range | Scale           | Border         |
| Tsukuba         | 284x288         | 0.15            | 16              | 18             |
| Venus           | 434x383         | 0.19            | 8               | 10             |
| Teddy           | 450x375         | 0.59            | 4               | 0              |
| Cones           | 450x375         | 0.59            | 4               | 0              |
Two different statistics have been defined to measure the quality of results, that is based on known ground truth disparity maps:

- Percentage of bad matching pixels, \(BadM\), between the computed disparity, \(D_c\), and the ground truth disparity, \(D_t\):
  \[
  BadM = \frac{1}{N} \sum_{(x,y)} |D_c(x,y) - D_t(x,y)| > \delta_{\text{error}}
  \]  
  where \(\delta_{\text{error}}\) is a disparity error tolerance. For the experiment in this work we use \(\delta_{\text{error}}=1.0\) since this coincides with [15], \(N\) is the total number of pixels.

- Root-Mean-Squared error (RMSE),
  \[
  RMSE = \left( \frac{1}{N} \sum_{(x,y)} |D_c(x,y) - D_t(x,y)|^2 \right)^{1/2}
  \]  
  the root mean squared error gives the error value the same dimensionality as the actual and computed values.

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 \(Dc\) and \(Dt\) where one of the images is a considered a true image and the other is the computed disparity. The PSNR is defined as:

\[
\text{PSNR} = 10 \log_{10} \left( \frac{\text{max}_{\text{pixel}}}{\text{RMSE}} \right)
\]

Here, \(\text{max}_{\text{pixel}}\) is the maximum pixel value of the image, when samples are represented using \(B\) bits per sample, maximum possible value of \(\text{max}_{\text{pixel}}\) is \((2^B-1)\).

In order to obtain the algorithm ranking, the four different images that are described in Table I\ref{Tab5:2} 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 \((B_{\delta})\), 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 \(\delta_{\text{error}}\). The first measure, for example, defined as:

\[
B_{\delta} = \frac{1}{|I|} \sum_{(x,y)\in I} |D_c(x,y) - D_t(x,y)| > \delta_{\text{error}}
\]

similarly, \(B_{ho}\) and \(B_{no}\) are defined for the half-occluded and near-occluded regions. Fig. 13, Table II show the application of these evaluations.

Table II compares the proposed algorithm with more than 25 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. The table shows the percentage of bad pixels, which are generated from our disparity, compared with the web side "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.

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):
  \[
  ADE = |D_c(x,y) - D_t(x,y)|
  \]
- The mean relative error (MRE) defined as:
  \[
  MRE = \frac{|D_c(x,y) - D_t(x,y)|}{D_t(x,y)}
  \]
- The percentage relative error (PRE): \[
  PRE = 100 \frac{|D_c(x,y) - D_t(x,y)|}{D_t(x,y)}
  \]

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 \(\mu\) 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.

\[
\mu = \frac{1}{N} \sum_{(x,y)} |D_c(x,y) - D_t(x,y)|
\]

It is nearly equal to Eq. 'ref'{Eq5:14} but here \(x, y\) run over the nonsingular points.

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

\[
\text{PRE} = 100 \frac{1}{N} \sum_{x,y} \frac{|D_c(x,y) - D_t(x,y)|}{|D_t(x,y)|}
\]

3) The deviation of the average error (DevAE):

\[
\text{DevAE} = \frac{1}{N} \sum_{(x,y)} |\mu - |D_c(x,y) - D_t(x,y)||
\]

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 \(D_t\) contain the spatial shift between corresponding pixels in the stereo images.

Therefore, the reconstruction of the right image defined as;

\[
RI_r(x,y) - D_c(x,y) = I_1(x,y)
\]

where \(x, y\) denote to the row and the column indices of the image.

The difference between the original image \(I_1\), and the reconstruction \(RI\), yield the error image \(EI_r\);

\[
EI_r(x,y) = I_1(x,y) - RI_r(x,y)
\]
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 $\xi$:

\[
\xi = \frac{\sum |E(x,y)|}{N \times K \times Rec}
\]

(27)

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

Finally, percentage of both total really wrong pixels ($\delta R$) and corrected pixels ($\eta$) 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.

Figure 14 shows a left image of the “Doll stereo pair”, the ground truth disparity, the computed disparity, and the 3-D surface reconstructed. Overall the experimental results the disparity estimated appear fairly accurate with in the majority of the surfaces. The comparison of the correspondence results of a line profile from reference disparity (TrueDisp) and the computed disparity (CompDisp) bottom-left. The plot shows the accurate and the smoothness of the computed result. That is because the phase-difference proposed method is optimally captures both local orientation and frequency information from the input image.

Figure 15(a) shows another example, “a cylinder view”, from these results we can say that the disparity estimates are obtained with subpixel accuracy, without requiring explicit subpixel signal reconstruction. The right plot in Fig.15(b) shows the amount of error in each pixel along the scan line, the maximum error is less than 0.2 pixel.

The intention of this work to demonstrated with the effects of slanted surface. Therefore the correctness of our algorithm immediately becomes evident when dealing with the stereo pairs shown in Fig. 16 (top). That is a left image of flat surface stereo pair rotated in 45° (depth increasing from right to left). The computed disparity map Fig. 16(c), show the disparities are linearly decreasing (i.e. depth are linearly increasing) from right to left (from bright to dark). Therefore, the results show how the proposed algorithm overcomes the problem of perspective foreshortening. 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. 17. The right plots showed the amount of error in each pixel along the scan line.

Figure 18 shows some reference data that used in the experimental results. Fig. 17 column (a) shows the left images from the stereo pairs, column (b) illustrates the reference disparity maps that so-called the ground truth disparity maps. The 3-D point clouds as well as the 3-D surface reconstruction are shown in columns (c) and (d) respectively.

SUMMARY AND CONCLUSION

We have presented a correction algorithm of local spatial frequency representation (combining magnitude and phase information) for disparity estimation, which takes into account not only the instability of phase but also surface perspective distortion (foreshortening). These properties are important to the use of phase information for avoid the incorrect disparity estimates. Instead of matching intensities directly, a Gabor scale-space expansion is used. In addition different magnitude criteria are used to detect “weak points” in the frequency domain and only reliable phase values remain for a robust estimation of resulting disparity. The method provides a foreshortening correction factor which verified to overcome the perspective distortion region, and demonstrates a novel solution to the phase-wraparound problem that has limited the application of other phase-based method. The presented test cases show that the performance of the proposed algorithm in terms of accuracy and density of the disparity estimates has greatly improved. Future work focuses on improvement the analysis by more complex images, and more fixed the validation of the derivatives.

REFERENCES

TABLE II. 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 SHOWS ONLY THE TEN TOP ALGORITHMS AND THE TWO LATEST ALGORITHMS IN THE EVALUATION

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Tsukuba</th>
<th>B_n</th>
<th>B_n</th>
<th>B_n</th>
<th>B_n</th>
<th>B_n</th>
<th>B_n</th>
<th>B_n</th>
<th>B_n</th>
<th>B_n</th>
</tr>
</thead>
<tbody>
<tr>
<td>[22]</td>
<td>5.98(1)</td>
<td>6.69(1)</td>
<td>13.4(1)</td>
<td>1.53(1)</td>
<td>2.19(1)</td>
<td>9.60(1)</td>
<td>8.43(4)</td>
<td>14.6(3)</td>
<td>19.3(3)</td>
<td>3.40(8)</td>
</tr>
<tr>
<td>[23]</td>
<td>6.21(5)</td>
<td>6.90(2)</td>
<td>12.9(2)</td>
<td>0.46(3)</td>
<td>0.85(5)</td>
<td>4.24(8)</td>
<td>6.29(6)</td>
<td>13.8(6)</td>
<td>16.0(9)</td>
<td>3.29(11)</td>
</tr>
<tr>
<td>[24]</td>
<td>6.10(4)</td>
<td>7.11(4)</td>
<td>14.6(2)</td>
<td>3.20(2)</td>
<td>3.80(2)</td>
<td>8.40(15)</td>
<td>14.7(22)</td>
<td>21.3(23)</td>
<td>24.4(21)</td>
<td>6.81(23)</td>
</tr>
<tr>
<td>[25]</td>
<td>5.81(2)</td>
<td>7.45(5)</td>
<td>15.8(11)</td>
<td>2.04(9)</td>
<td>3.45(20)</td>
<td>17.7(26)</td>
<td>17.9(1)</td>
<td>21.1(18)</td>
<td>31.3(18)</td>
<td>5.97(13)</td>
</tr>
<tr>
<td>[26]</td>
<td>8.26(10)</td>
<td>10.4(10)</td>
<td>23.0(12)</td>
<td>2.83(1)</td>
<td>3.89(3)</td>
<td>29.1(35)</td>
<td>16.5(26)</td>
<td>25.1(26)</td>
<td>34.0(31)</td>
<td>9.15(23)</td>
</tr>
<tr>
<td>[27]</td>
<td>18.7(25)</td>
<td>19.1(19)</td>
<td>15.8(6)</td>
<td>0.83(1)</td>
<td>1.34(8)</td>
<td>3.84(3)</td>
<td>5.77(2)</td>
<td>11.4(4)</td>
<td>13.4(1)</td>
<td>4.80(8)</td>
</tr>
<tr>
<td>[28]</td>
<td>11.2(11)</td>
<td>13.4(12)</td>
<td>23.9(8)</td>
<td>7.81(1)</td>
<td>9.28(3)</td>
<td>27.9(11)</td>
<td>30.4(12)</td>
<td>28.5(11)</td>
<td>32.7(9)</td>
<td>15.9(6)</td>
</tr>
<tr>
<td>[29]</td>
<td>18.1(19)</td>
<td>18.8(18)</td>
<td>18.5(7)</td>
<td>1.32(1)</td>
<td>1.09(1)</td>
<td>8.27(14)</td>
<td>10.1(16)</td>
<td>16.3(14)</td>
<td>23.5(19)</td>
<td>6.17(16)</td>
</tr>
<tr>
<td>[30]</td>
<td>19.0(21)</td>
<td>20.7(25)</td>
<td>17.5(14)</td>
<td>3.98(1)</td>
<td>5.20(25)</td>
<td>15.9(25)</td>
<td>14.7(4)</td>
<td>21.6(24)</td>
<td>25.1(22)</td>
<td>15.7(17)</td>
</tr>
<tr>
<td>[26]</td>
<td>19.1(22)</td>
<td>19.5(26)</td>
<td>20.0(1)</td>
<td>1.67(1)</td>
<td>2.05(3)</td>
<td>4.63(1)</td>
<td>9.28(1)</td>
<td>15.2(1)</td>
<td>20.2(9)</td>
<td>5.62(12)</td>
</tr>
<tr>
<td>[31]</td>
<td>24.3(30)</td>
<td>26.1(34)</td>
<td>44.8(8)</td>
<td>9.65(33)</td>
<td>11.0(33)</td>
<td>42.6(36)</td>
<td>18.2(1)</td>
<td>25.8(11)</td>
<td>42.8(34)</td>
<td>11.8(31)</td>
</tr>
<tr>
<td>[15]</td>
<td>28.5(26)</td>
<td>30.0(26)</td>
<td>38.1(24)</td>
<td>4.98(20)</td>
<td>6.43(26)</td>
<td>14.5(24)</td>
<td>19.8(26)</td>
<td>27.8(20)</td>
<td>38.2(20)</td>
<td>13.8(30)</td>
</tr>
</tbody>
</table>

Fig. 14. Comparison of the correspondence line profile from reference disparity (true) and the computed disparity (comp). The error map contain a bad pixels in (in magenta), correct pixels in gray level from 32 to 64 color map.
Figure 15. 3-D surface reconstruction for cylinder stereo pair images.

Figure 16. 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.
Figure 17. Ground truth (TrueDisp) compare against the computed disparity maps (CompDisp), for the center scanline of the flat surface at the various rotation angle; 0°, 20° and 45° and disparity range (1...6) pixels.

Figure 18. 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.