STEREO PAIR MATCHING OF ARCHAEOLOGICAL SCENES USING PHASE DOMAIN METHODS

Manthos Alifragis and Costas S. Tzafestas

National Technical University of Athens, School of Electrical and Computer Engineering
Division of Signals, Control, and Robotics, Zographou Campus, 15773 Athens, Greece

Abstract: This paper conducts an experimental study on the application of some recent theories of image preprocessing and analysis in the frequency domain, particularly the phase congruency and monogenic filtering methods. Our goal was to examine the performance of such methods in a stereo matching problem setting, with photos of complicated scenes. Two objects were used: a scene of an ancient Greek temple of Acropolis and the outside scene of the gate of an ancient theatre. Due to the complex structure of the photographed object, classic techniques used for feature detection and matching give poor results. The phase-domain approach followed in this paper is based on the phase-congruency method for feature extraction, together with monogenic filtering and a new correlation measure in the frequency domain for image correspondence and stereo matching. Comparative results show that the three-dimensional models of the scene computed when applying these phase domain methods are much more detailed and consistent as compared to the models obtained when using classic approaches or the SIFT based techniques, which give poor depth representation and less accurate metric information.

1 INTRODUCTION

The problem of stereo matching and depth estimation have become in recent years the focus of considerable research in the field of computer vision. Reliable edge or feature detection techniques constitute the precursors of three dimensional structure or scene reconstruction methods. Throughout the years, there has been a significant progress in the development of image correspondence analysis and feature detection methods. As far as feature (edge or corner) detection is concerned, the traditional approach endorsed in most of the applications, is the one applying gradient-based methods, such as those developed by Canny (Canny, 1986), and Marr & Hildreth (Marr and Hildreth, 1980). These methods have the drawback of sensitivity in image illumination, contrast, blurring and magnification. Another disadvantage when using these methods is the non-automatic determination of the appropriate thresholds for feature detection. More recently, Fleck (Fleck, 1992) used an a-priori knowledge of the noise characteristics of the camera, in order to set feature detection thresholds. A remarkable study on the detection of image features invariant to image scale and rotation has been

made by Lowe (Lowe, 2004). This specific approach has been named Scale Invariant Feature Transform (SIFT). This approach imposes a local image descriptor which is highly distinctive and invariant to image scale-space variations and changes in illuminations or 3D viewpoint, too.

In the work presented in this paper, we made use of Fourier transformations of the images, and Gabor filtering, together with the phase congruency approach proposed by Kovesi (Kovesi, 1999). This approach utilizes the local frequency spread and uses this information to weigh the phase congruency measure of the image. Concerning image correspondence, we extend the approach proposed in (Kovesi, 1995), using monogenic filtering together with a new correlation measure in the frequency domain. This approach was enriched by a normalized expression of the correlation measure, as well as by the additional information of line detection results and an approximately known camera motion. The combination of the above methods led us to the development and implementation of a filter in the frequency domain, which was experimentally applied in a stereo matching problem.

The rest of the paper is organized as follows: Sec-

tion 2.1 presents the computational approach applied in the frequency domain for image processing and edge detection. Section 2.2 presents the new stereo-matching approach that is based on the application of monogenic filtering. Section 2.3 makes a brief discussion on camera calibration issues, also presenting the image rectification methods that were used in our experiments. Finally, an evaluation of all the above mentioned methods in comparison with the traditional approaches as applied to the problem of stereo matching and depth estimation for 3D scene reconstruction, is presented in Section 3.

2 PHASE DOMAIN METHODS

2.1 Feature Extraction using Phase Congruency

The first phase in a stereo analysis problem is the extraction of useful information from the images. This phase consists of detecting edges or other geometrical structures of interest on the images. The detection of sporadic or texture points is avoided. The traditional approach to the edge detection problem is based on intensity information (Canny, 1986). It is known, however, that these kind of intensity based methods are very sensitive to random lighting of the photographs, to optical characteristics of the camera and are mostly dependent on the different levels of thresholding.

The development of local energy models to detect image features, such as edges or curves, under random lighting conditions is presented in (Morrone and Owens, 1987). According to these models and through two dimensional fourier transformation of an image, local frequency and phase information is obtained. The phase congruency definition by Morrone and Owens (Morrone and Owens, 1987) is given by the Fourier series expansion of an one dimensional signal at some location x as:

$$PC(x) = max_{\overline{\varphi}(x) \in [0,2\pi]} \frac{\sum_n A_n cos(\varphi_n(x) - \overline{\varphi}(x))}{\sum_n A_n} \quad (1)$$

The quantity φ represents the local phase angle of the Fourier transformation at the specific image point. The quantity $\overline{\varphi}$ which maximizes Eq. (1) is the mean phase angle of all Fourier components at a local neighborhood of the image. In the approach followed in this paper, we use complex Gabor functions (sine and cosine functions modulated over gaussian). In order to measure, in a specific location of the signal, the amplitude and the phase, we can apply two linear phase filters in quadrature as Gabor complex filters,

for a specific scale and frequency. The Gabor filter is composed of two main components, the complex sinusoidal carrier, and a gaussian envelope. Alternatively, log-Gabor filters can be applied as proposed by to Field (Field, 1987). The log-Gabor filters have gaussian transfer function and allow the construction of large bandwidth filters with odd symmetry and DC component equal to zero. The zero DC component cannot be kept in Gabor filters with bandwidth greater than one octave. The log Gabor function has a frequency response described by:

$$G(f) = e^{-[log(f/f_0)]^2/2[log(\sigma/f_0)]^2} \quad (2)$$

The frequency response of a log-Gabor filter is a Gaussian on a log frequency axis. The f_0 defines the centre frequency of the sinusoid and represent the scaling factor of the filter. In addition, the σ is a scaling factor of the bandwidth. In order to maintain constant shape ratio filters, the ratio of σ/f_0 should be maintained constant. The first step of this analysis consisted of the convolution of the signal with each quadrature part of the filter. Let I be the signal, and M_n^e, M_n^o be the even (cosine) and odd (sine) symmetry waveforms, respectively, at scale n. The response vector is then consisted of the responses of each quadrature pair of filters. It is written as:

$$(e_n(x), o_n(x)) = (I(x) * M_n^e, I(x) * M_n^o) \quad (3)$$

where $e_n(x)$ and $o_n(x)$ are the real and imaginary parts of the frequency Fourier component. The amplitude of the transformed signal at a specific scale of the gabor filter is given by:

$$A_n(x) = \sqrt{e_n(x)^2 + o_n(x)^2} \quad (4)$$

The phase is given by:

$$\Phi_n(x) = arctan(e_n(x), o_n(x)) \quad (5)$$

at each point x in a signal. The *arctan* is the four quadrant inverse tangent.

A response vector was thus formed for each scale of the filter. The array of all these vectors represents the localized information of a signal. The above expression (1) of the phase congruency measure has poor localization on blurred features. This measure is also sensitive to noise. These issues led Kovesi in (Kovesi, 1999) to develop a new measure for phase congruency and to extend the one-dimensional filters described previously into two dimensions. This new measure includes the additional sum of the log-Gabor filters response amplitudes multiplied in the frequency domain with some spreading function over all orientations and scales at a specific location in the image.

Gaussian functions were also used as spreading functions across the filter perpendicular to its orientation, according to Kovesi (Kovesi, 1999). This resulted in preserving the phase information unaffected, due to the fact that any signal convolved with a gaussian function has its amplitude components modulated, while the phase remains unaltered. Towards this direction, the two dimensional filters used in the frequency domain are Gaussian with geometrically increasing center frequencies and bandwidths. Their transfer function is:

$$G(\theta) = e^{-\frac{(\theta - \theta_0)^2}{2\sigma_\theta^2}} \qquad (6)$$

where θ_0 is the orientation angle of the filter, and σ_θ is the standard deviation of the gaussian spreading function in the angular direction.

The equation for 2-D phase congruency for two-dimensional signals analysis, like images, can be obtained as follows:

$$PC_2 = \frac{\sum_o \sum_n W_o(x) \lfloor A_{no} \triangle \phi_{no}(x) - T \rfloor}{\sum_o \sum_n A_{no}(x) + \varepsilon} \qquad (7)$$

at each two-dimensional image location x. The $\triangle \phi_{no}(x) = \cos(\phi_{no}(x) - \overline{\phi}(x)) - |\sin(\phi_{no}(x) - \overline{\phi}(x))|$, and o, n refer to the filter's orientation and scale, respectively. It must be also noted that when the amount between the symbols $\lfloor . \rfloor$ is non-positive, then the outcome becomes zero. The numerator of the above fraction represents the total energy of the 2D signal at a local point of the image. This amount of energy is an approximation of the local energy function defined for an analytical signal, according to Venkatesh and Owens (Venkatesh and Owens, 1989). The term $W_o(x)$ weighs the frequency spread. Kovesi made use of $W_o(x)$ as a component to cope with the lack of reliability of phase congruency measures in image areas with less frequency spread (e.g. smoothed images). The role of ε is to avoid division by zero. Finally, only the values which are above a threshold T (the expected influence of noise) are used to calculate the final result. The appropriate threshold T for the noise is set experimentally, according to the response of the smallest scale filter on each image.

It must be also noted here that the type of an image feature detected, such as a line or a corner corresponding to maximum phase congruency value, needs to be classified accordingly. Towards this end, the phase congruency feature maps were calculated, according to (Kovesi,).

2.2 Feature Matching on the Phase Domain

The image correspondence problem has been excessively studied as a fundamental problem of low-level computer vision. In order to track correspondent points through images, intensity correlation processes can be applied. However, the apparent complexity of the images used in our experiments, dictated the use of a different approach to address this issue. In the work presented in this paper, we employ an approach based on two-dimensional analytic signal theory and monogenic signal theory, inspired by the work of Sommer and Ferlsberg in (Felsberg and Sommer, 2001).

The two dimensional analytic signal is based on a two dimensional generalization of Hilbert transformation, also known as Rietz transformation. The expression of the Rietz-transformed signal $F(\mathbf{u})$ in the frequency domain:

$$F_R(\mathbf{u}) = i\frac{\mathbf{u}}{|\mathbf{u}|}F(\mathbf{u})$$

where $\mathbf{u}$ the two dimensional frequency vector (u_1, u_2), and $|\mathbf{u}| = \sqrt{u_1^2 + u_2^2}$.

The Fourier transformation of each image I_F was firstly computed. The next stage was the introduction of a log-Gabor filter (see Eq.(2)), which contributes to the construction of bandpass expressions of the signal $\mathbf{F}_R$ in the frequency domain:

$$\mathbf{H}_{R_s} = \mathbf{F}_R G(f) \qquad (8)$$

where $\mathbf{H}_{R_s} = \begin{pmatrix} H_{R_s}^1 & H_{R_s}^2 \end{pmatrix}^T$. The above image filters are applied in the frequency domain, and after the application of an inverse Fourier transformation, the real part of the consequent signals is obtained as follows:

$$I_{F_s} = Re\left[F^{-1}\left\{I_F G(f)\right\}\right] \qquad (9)$$

$$H_{F_s}^1 = Re\left[F^{-1}\left\{I_F H_{R_s}^1\right\}\right] \qquad (10)$$

$$H_{F_s}^2 = Re\left[F^{-1}\left\{I_F H_{R_s}^2\right\}\right] \qquad (11)$$

This leads to a generalized complex 2D analytic signal expression, which has as real part the signal I_{F_s} (9) and complex part the mathematical expression of its Rietz transformation, according to Ferlsberg and Sommer (Felsberg and Sommer, 2001). Therefore, the complex part of the 2D analytic signal consists of two signals, the signal $H_{F_s}^1$ (10) and the signal $H_{F_s}^2$ (11). Consequently, at each point of the image (x, y), at a specific scale and orientation, we have a 3D vector $\mathbf{x}(x, y)$ consisting of the three above signals (9), (10) and (11). In addition, the measure of the amplitude of the energy is given by:

$$A = \sqrt{(H_{F_s}^1)^2 + (H_{F_s}^2)^2 + (I_{F_s})^2} \qquad (12)$$

Given two locations (x_1, y_1) and (x_2, y_2) in the first and second image (Im_1 and Im_2) respectively, the correlation measure is then given by:

$$C_{12}(\mathbf{x}_1^{Im_1}, \mathbf{x}_2^{Im_2}) =$$
$$\frac{\sum_{m=-k}^{+k} \sum_{n=-l}^{+l} \mathbf{x}_1^{Im_1}(x_1 + m, y_1 + n)}{\sum_{m=-k}^{+k} \sum_{n=-l}^{+l} A^{Im_1}(x_1 + m, y_1 + n)}$$
$$\frac{\mathbf{x}_2^{Im_2}(x_2 + m, y_2 + n)}{A^{Im_2}(x_2 + m, y_2 + n)} \qquad (13)$$

where $\mathbf{x}_1^{Im_1}$ and $\mathbf{x}_1^{Im_2}$ are the three dimensional monogenic filter responses vectors of candidate points (x_1, y_1) and (x_2, y_2) for matching. The correlation measure, in (13) was computed by the dot product of above vectors with window $-l$ to l and $-k$ to k in the two-dimensional image plane. In addition, this measure was normalized by the sum of the amplitude responses, according to (12). The matching is considered successful for the pairs of points where the above correlation measure is maximized (i.e. $argmax|C_{12}(\mathbf{x}_1^{Im_1}, \mathbf{x}_2^{Im_2})|$).

2.3 Camera Calibration and Image Rectification

In the first phase of our experiments, the camera was calibrated using the algorithm presented by Zhang in (Zhang, 1999) with planar patterns. Regarding the camera model, we assumed, that there is no skew. The focal length per distance unit for the two image plane directions is represented by α_x and α_y values. The values p_x and p_y represent the coordinates of the principal points in x and y direction accordingly. Following these assumptions, the matrix of the camera intrinsic parameters, takes the following the form:

$$K = \begin{pmatrix} \alpha_x & 0 & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{pmatrix} \qquad (14)$$

The process of image rectification simplifies the matching problem by transforming the whole search area for each image from 2D to 1D. Therefore, the epipolar lines become parallel and coincide with the scan lines used to find matching pairs. The rectification transformation used to place the epipolar lines in parallel, was based on the above assumptions concerning the camera intrinsic parameters, according to Hartley (Hartley, 1997) and Koch et al. (Koch et al., 1998).

Figure 1: Successive photos, in grayscale, of a temple in Acropolis and the gate of an ancient theatre, captured with a camera moving in a straight line.

3 IMPLEMENTATION - RESULTS

3.1 Feature Detection - Matching

The camera that we used for the experiments was a Nikon D70s 18-74mm. Successive photographs of the side views of the two subjects (namely, a temple in Acropolis of Athens and the outside scene of an ancient theatres gate) were used as experimental data, while the camera was sliding in an almost straight line. A short displacement was used to avoid having large occluded areas. All the images were transformed in grayscale, and color information was not used for depth estimation. The photos captured and used for the experiments are shown in Fig.1.

The initial phase in a stereo matching process is feature detection for each image frame. The three main approaches for feature detection that were experimentally tested (in comparison) are presented in the sequel. The first approach consisted of applying classic filters based on image intesity gradient computations such as the Canny edge detector. In the second approach a scale-space representation of the image was utilized in order to extract features (keypoints) that can be repeatedly detected under slightly different views or any change in image scale, rotation, or illumination conditions. The candidate keypoints have been detected according to Lowe (Lowe, 2004) using scale-space extrema in the difference of Gaussian function convolved with the image. In the third approach, edge detection on the rectified images, was performed based on phase congruency filtering. In this case, edges were detected on the images through the calculation of the maximum value of the moments of phase congruency covariance, as it has already been described. From the group of detected edges, we chose those with length greater than

a selected threshold. The feature detection results for both approaches are shown in Fig.2.

For the task of edge detection based on phase congruency concepts, log-Gabor functions was used, with Gaussian transfer functions on a logarithmic frequency scale. This filter was applied in six orientations and at four scales, with a constant one octave bandwidth, according to Equations (7), (2), and following analysis presented in (Field, 1987). By observing the first row in Fig. 2 one can conclude that the application of traditional techniques, like Canny filtering, for such a complicated scene results in poor localization of the detected edges, as compared to the results obtained by applying phase congruency methods shown in the third row of Fig. 2. The second row in Fig. 2 shows the detected canditate SIFT keypoints at a specific level of the constructed scale space pyramid. In this case, it is evident that the use of difference of Gaussian operator, which is based on gradient measurements, emphasizes edge features, even those features with low contrast. This kind of low contrast features will be excluded from SIFT features as being non distinctive. The apparent complexity of the scene resulted in a large number of these features, therefore, the matching process will be based on less candidate points.

The next stage in depth estimation is the process of matching corresponding points between successive images, which is known as an ill-conditioned problems in low-level vision. The quality of the solution of the matching problem has a direct impact on the quality of the scene reconstruction. The matching process was again performed based on three methods, in comparison: (a) The first approach used Canny filtering for feature detection and a typical intensity-based correlation method for the matching process. (b) The second approach consisted of SIFT keypoints detection, based on image gradient amplitude and orientation measurements, the construction of invariant keypoint descriptors for each image of the stereo pair and, finally, the matching process which was based on these descriptors correspondence, through Euclidean distance measurements. (c) Finally, the third approach followed in this paper was based on the application of monogenic filters, as has been described in section 2.2. The specific and unique direction of camera motion indicated the direction on which the candidates for feature matching were moved on the image plane. The process of rectification also locates corresponding points on the same line. Based on these remarks the search area for image correspondences was radically reduced, resulting effectively in the calculation of much more reliable matching points.

During our experiments, we observed that the

matching of sporadic points created many problems, especially when an intensity-based correlation method was applied. This occurs because the photos were outdoor, very complicated and had been taken under random lighting and illumination conditions. This means that the intensity values of specific points include a lot of uncertainty. Variations in shading (in one or more photos), repeated patterns on the images and a uniform texture, all result in very close intensity values for certain pixel neighborhoods, leading to a large number of candidate points for matching. The first attempt to overcome the uncertainty problems was based on the use of a correlation measure for whole geometric primitives like lines, excluding from the correlation process sporadic points.

The prior knowledge of the camera motion was used in order to look for probable corresponding lines in the opposite direction of which the camera was moving. The search area was basically restricted on a horizontal axis on the image, due to the known horizontal motion of the camera. The search window was chosen to have its (horizontal dimension) width almost equal to the half of the image width, and its height (vertical dimension) equal to a few pixels. However, line characteristics like length and direction may present considerable deviation between corresponding images. It is, for instance, possible for a detected line on one image to break into two or more parts on the other image. Therefore, we decided, instead, to use point correlation techniques on these candidate lines. Such an approach improved indeed the obtained results. The confirmation of the matching validity in the neighborhood of each point, was achieved by the implementation of classic relaxation methods, as in (Faugeras, 1993).

3.2 3D Reconstruction Results

Our main goal in this study was to evaluate the performance of the phase domain methods, in comparison to the classical (intensity based) filtering techniques and the SIFT local keypoint descriptors in a stereo matching and 3D calibrated reconstruction problem. The feature matching process led to the acquisition of matching pairs in the two images. Hence, the estimation of the fundamental matrix becomes feasible. The calibration matrix was recovered by implementing the Zhang's method as briefly described in Section 2.3. Consequently, the projection matrices of the camera were computed in both configurations through the essential matrix according to the known relation introduced by Hartley and Zisserman (Hartley and Zisserman, 2000). The metric information of the scene was recovered using the camera calibration matrix.
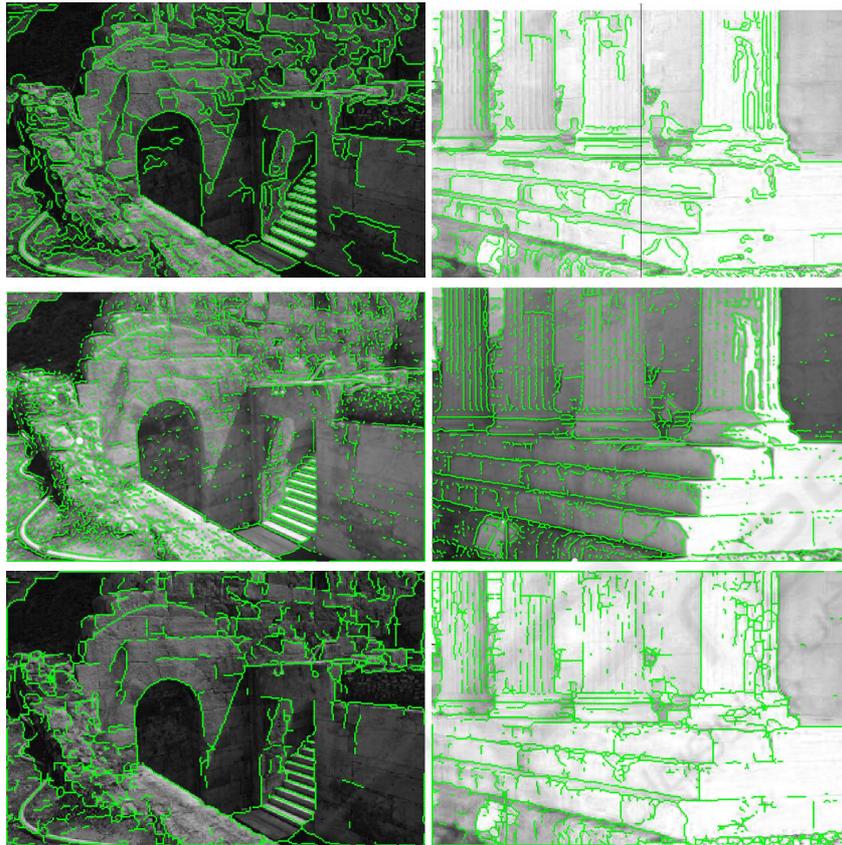
Figure 2: Edge detection results for the two photos. First Row: Application of Canny filter. Second Row: Use of Gaussian filter. Third Row: Application of phase congruency methods.

Depth estimation for each matching pair of points was performed by point triangulation, and was further refined by minimizing the reprojection error using the Levenberg-Marquadt method.

Comparative results for the three different methods are presented in Figures 3(a) and 3(b). The first row to both of these subfigures presents the results obtained when applying classic methods both for feature detection and point matching, that is, Canny edge detection and intensity-based correlation techniques. For the results depicted in the second row, the difference of Gaussian filter was applied for SIFT keypoints detection. The matching process for image points approached by constructing local image descriptors. Local image region descriptors assign the gradient magnitude and orientation to each keypoint according to (Lowe, 2004). The best candidate match was found by identifying of that point, on the other image of the stereo pair, with minimum Euclidean distance for the invariant keypoint descriptor vector. In the third row, the results are obtained using both the phase congruency method for edge detection and the monogenic filtering approach for point correlation.

The results presented in Fig. 3(a) and 3(b), are also organized in three columns. The first column of each subfigure shows the 3D scene reconstruction results, for the three methods mentioned above, while the second column depicts the same reconstructed scenes but rotated by a small angle (approximately 20 degrees, to better illustrate the estimated scene depth). Finally, the third column presents a colored illustration of the scene depth, with red being the color of the nearest points and blue the color of the most distant ones (with linear color to distance variation).

By observing these experimental results, it can be seen that a more dense and accurate representation of the scene structure is obtained when all computations are conducted in the phase domain (third row in Fig. 3(a) and 3(b)). In addition, we observe that a better variance of the depth values is obtained when the matching process is implemented in the phase domain. This observation becomes more evident when the reconstructed scene is rotated as shown in the second column of Figures 3(a) and 3(b). The classic methods give results that are evidently non-satisfactory in this case, regarding the structure and

26

(a) Ancient Temple Reconstruction

(b) Gate of Ancient Theatre
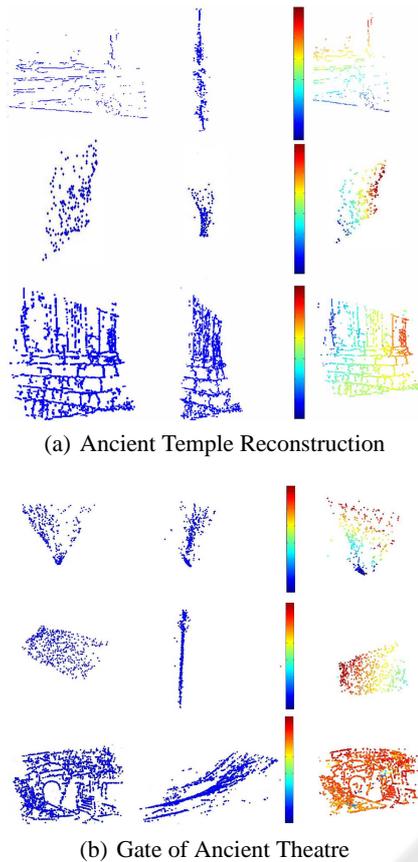
Figure 3: Scene reconstruction results obtained from two experimental stereo pairs. The organization of the results in rows and columns is the same to both of subfigures. First row: classic techniques applied using only the intensity values of each image. Second row: scene reconstruction results using gradient measurementsfor edge detection and SIFT keypoints matching for scene reconstruction. Third row: computations conducted solely in the phase domain. First Column: scene reconstruction results using different edge detection and matching methods. Second column: reconstruction results rotated by a small angle. Third column: colored illustration of the scene depth, with red being the color of the nearest and blue the color of the most distant points.

depth estimation of the scene. The computation of the three dimensional points by an intensity-based correlation method (first row results in figure 3(a) and 3(b)), in this case, leads to a structure where all points are almost co-planar with a very vague reconstructed scene structure and a lot of outliers (for the results depicted in the first row). Furthermore, the results of the second row in both of the subfigures indicate that the use of SIFT keypoints descriptors is also not an appropriate procedure for stereo matching in such cases of images with large scale of complexity. The computation of keypoint descriptors is based on gradi-

ents measurements in the image in different levels of Gaussian blur. That kind of measurements can detect features with poorly defined peaks in the difference-of-Gaussian function, which will be rejected from keypoint descriptor computation. This results in the detection of sporadic keypoints that do not follow a specific structure. The same conclusion can be confirmed by the very poor depth variation without a clear sense of the scene structure, in the second column of the second row results for both figures 3(a) and 3(b).

4 CONCLUSIONS AND FUTURE WORK

This paper presents an approach for depth estimation and scene reconstruction using phase domain methods, based on concepts that involve local representation of image features. We implement recent ideas of local energy models for each stage of the 3D reconstruction process, comprising mainly the tasks of feature detection and image correspondence. More specifically, for the task of detecting edges as image features, we applied the phase congruency method, introduced by Kovesi (Kovesi, 1999), while for the image correspondence task we implemented a new version of a correlation measure based on monogenic filtering. An appropriate normalization was performed, based on localized amplitude responses of log-Gabor signals and a prior-knowledge of the camera motion, in order to enrich the mathematical expression of the new correlation measure. Experimental results showed that feature estimation in the frequency domain remains invariant to changes in the lighting conditions of the scene. It was concluded that the proposed approach leads to more reliable results, producing a more accurate metric information of the scene and a more dense structure regarding the outcome of the 3D scene reconstruction process. The good behavior of such phase-domain models was confirmed, which seem to present a choice of preference for the task of image-based 3D reconstruction of complex sceneries, such as the archaeological site or an outside scene used in the experimental study of this work. In the future we plan to extend the approach presented in this paper to extract dense disparity maps from multiple camera views, integrated within probabilistic frameworks.

ACKNOWLEDGEMENTS

REFERENCES

Canny, F. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8:112–131.

Faugeras, O. (1993). *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, Cambridge, Massachussets.

Felsberg, M. and Sommer, G. (Dec 2001). The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12).

Field, D. J. (Dec. 1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of The Optical Society of America A*, 4(12):2379–2394.

Fleck, M. M. (March 1992). Multiple widths yield reliable finite differences. *IEEE T-PAMI*, 14(3):337–345.

Hartley, R. (1997). In defence of the eight-point algorithm. *IEEE T-PAMI*, 19(6):580–593.

Hartley, R. and Zisserman, A. (2000). *Multiview Geometry in Computer Vision*. Cambridge University Press.

Koch, R., Pollefeys, M., and Gool, L. V. (1998). Automatic 3d model acquisition from uncalibrated image. In *Proceedings Computer Graphics International*, pages 597–604, Hannover.

Kovesi, P. D. Matlab and octave functions for computer vision and image processing. Available from: http://www.csse.uwa.edu.au/ pk/research/matlabfns.

Kovesi, P. D. (1995). Image correlation from local frequency information. In *The Australian Pattern Recognition Society Conference: DICTA'95*, pages 336–341, Brisbane.

Kovesi, P. D. (1999). Image features from phase congruency. *Videre: A Journal of Computer Vision Research, MIT Press*.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Marr, D. and Hildreth, E. C. (1980). Theory of edge detection. In *Proceedings of the Royal Society, London B*, pages 187–217.

Morrone, M. C. and Owens, R. A. (1987). Feature detection from local energy. *Pattern Recognition Letters*, 6:303–313.

Venkatesh, S. and Owens, R. (1989). An energy feature detection scheme. In *International Conference on Image Processing*, pages 553–557, Singapore.

Zhang, Z. (1999). A flexible new technique for camera calibration. In *International Conference on Computer Vision*, Kerkyra, Greece.