

Research Article

A Multi-Model Stereo Similarity Function Based on Monogenic Signal Analysis in Poisson Scale Space

Jinjun Li,^{1,2} Hong Zhao,¹ Chengying Shi,² and Xiang Zhou¹

¹ State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

² Xi'an Research Institute of Hi-Tech, Hongqing Town, Xi'an, Shaanxi 710025, China

Correspondence should be addressed to Jinjun Li, jinjun_lee@stu.xjtu.edu.cn

Received 28 July 2010; Revised 15 November 2010; Accepted 20 March 2011

Academic Editor: Katica R. (Stevanovic) Hedrih

Copyright © 2011 Jinjun Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A stereo similarity function based on local multi-model monogenic image feature descriptors (LMFD) is proposed to match interest points and estimate disparity map for stereo images. Local multi-model monogenic image features include local orientation and instantaneous phase of the gray monogenic signal, local color phase of the color monogenic signal, and local mean colors in the multiscale color monogenic signal framework. The gray monogenic signal, which is the extension of analytic signal to gray level image using Dirac operator and Laplace equation, consists of local amplitude, local orientation, and instantaneous phase of 2D image signal. The color monogenic signal is the extension of monogenic signal to color image based on Clifford algebras. The local color phase can be estimated by computing geometric product between the color monogenic signal and a unit reference vector in RGB color space. Experiment results on the synthetic and natural stereo images show the performance of the proposed approach.

1. Introduction

Shape and motion estimation from stereo images has been one of the core challenges in computer vision for decades. The robust and accurate computation of stereo depth is an important problem for many visual tasks such as machine vision, virtual reality, robot navigation, simultaneous localization and mapping, depth measurements, and 3D environment reconstruction. Most of conventional approaches, such as intensity-based or correlation-based matching, feature-based matching, and matching function optimization techniques, estimate the disparity only based on local intensity and feature between stereo images so that the results may be susceptible to level shift, scaling, rotation, and noise [1, 2].

To overcome these drawbacks, we propose a new method for establishing spatial correspondences between a pair of color images. Unlike classical stereo-matching method

based on brightness constancy assumption and phase congruency constraint, we match feature points and estimate disparity map between stereo images based on a new local multi-modal monogenic image feature descriptors in the Color Monogenic Signal framework [3, 4]. We firstly introduce the monogenic signal [5] of 2D gray level image using Dirac operator and Laplace equation and extract local amplitude, local orientation, and instantaneous phase information in multiscale space [6]. At the same time, the 2D monogenic signal is extended to color image, and the color monogenic signal is introduced based on Clifford algebras. The local color phase is estimated by computing geometric product between the color monogenic signal and a unit reference vector in RGB color space with Clifford algebras [3, 4]. Then we focus on defining new local multi-model monogenic image feature descriptors which contain local geometric (local orientation), structure (instantaneous phase), and color (local color phase and color values) information in the Color Monogenic Signal framework. Based on the proposed image feature descriptors, a stereo similarity function between two primitives is also defined to solve stereo correspondence problem. Finally, we test the performance of the proposed approach on the synthetic and natural stereo images, and experiment results are given in detail.

2. Monogenic Signal Analysis in Poisson Scale Space

2.1. Modeling 2D Image Signal

Based on the results of Fourier theory and functional analysis, we assume that each 2D signal $f \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ can be locally modeled by a superposition of arbitrarily orientated one-dimensional cosine waves [6]:

$$(p_s * f)(x, y) = \sum_{v=1}^n a_v(x, y, s) \cos(\langle x, y, \bar{o}_v(x, y, s) \rangle + \phi_v(x, y, s)) \quad (2.1)$$

with $*$ as the convolution operator and the orientation $\bar{o}_v(x, y, s) = [\cos \theta_v(x, y, s), \sin \theta_v(x, y, s)]^T$. Note that each cosine wave is determined with the same amplitude and phase information. The Poisson convolution kernel reads

$$p_s(x, y) = \frac{s}{2\pi(s^2 + x^2 + y^2)^{3/2}}. \quad (2.2)$$

For a certain scale space parameters $s \in \mathbb{R}_+$, the Poisson kernel acts as a low pass filter on the original signal f . The Poisson scale space is naturally related to the generalized Hilbert transform by the Cauchy kernel. To filter a frequency interval of interest, the difference of Poisson (DoP) kernel will be used in practice:

$$p_{s_f, s_c}(x, y) = p_{s_f}(x, y) - p_{s_c}(x, y) \quad (2.3)$$

with $s_c > s_f > 0$ and s_c as the coarse scale parameter and s_f as the fine scale parameter. The filtered signal is defined by convolution with the difference of Poisson kernel which will be used to analyze the original with the DoP operator to consider only a small passband of the

original signal spectrum. Without loss of generality the signal model in (2.1) degrades locally at the origin $(x, y) = (0, 0)$ of a local coordinate system to

$$f_p(x, y, s) = \sum_{v=1}^n a(x, y, s) \cos \phi(x, y, s). \quad (2.4)$$

In case of image analysis lines, edges, junctions, and corners can be models in this way. The signal processing task is now to determine the local amplitude $a(x, y, s)$, the local orientation $\theta_v(x, y, s)$, and the local phase $\phi(x, y, s)$ for a certain scale space parameter s and a certain location (x, y) . This problem has been already solved for one-dimensional signals by the classical analysis [7] by means of the Hilbert transform [8] and for intrinsically one-dimensional signals [9] by the two-dimensional monogenic signal by means of the generalized first-order Hilbert transforms [10].

2.2. The Analytic Signal

Let $s: \mathbb{R} \rightarrow \mathbb{R}$ be a real-valued signal, and let $f: \mathbb{R} \rightarrow \mathbb{R}_{2,0}$ [3] be a vector-valued signal such that $f(x) = s(x)e_2$ [4]. The purpose is to construct a function fulfilling the Dirac equations whose real part is real-valued signal. It is equivalent to find the solution of a boundary value problem of the second kind (a Neumann problem):

$$\begin{aligned} \Delta u &= \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad \text{if } y > 0, \\ e_2 \frac{\partial u}{\partial y} &= f(x), \quad \text{if } y = 0, \end{aligned} \quad (2.5)$$

with $D = e_1(\partial/\partial x) + e_2(\partial/\partial y)$ and $\Delta = D^2$.

The first equation in (2.5) is the 2D Laplace equation restricted to the open domain $y > 0$. The second equation is called the boundary condition and the basis vector e_2 is coherent with the embedding of complex functions as fields (the real part is embedded as the e_2 -component). Using the fundamental solution of the 2D Laplace equation, the solution of the problem leads to

$$f_A(x, y) = p_1 * f(x, y) + p_1 * h_1 * f(x, y), \quad (2.6)$$

where $p_1 = y/\pi(x^2 + y^2)$ is the 1D-Poisson kernel and $h_1 = (1/\pi x)e_{12}$ is the Hilbert kernel. The variable y is a scale parameter, and, taking it equal to zero, the classical analytic signal can be obtained.

2.3. Instantaneous Phase of the Monogenic Signal

Following the previous construction of the analytic signal, Michael Felsberg and Gerald Sommer has proposed an extension to 2D signals and defined a monogenic signal which is the combination of a gray image with its Riesz transform [5]. Let $s: \mathbb{R}^2 \rightarrow \mathbb{R}$ be a real-valued

signal and $f : \mathbb{R}^2 \rightarrow \mathbb{R}_{3,0}$ a vector-valued signal, and $\{e_i\}$ ($i = 1, 2, 3$) is the orthonormal basis of \mathbb{R}^3 such that $f(x, y) = f_3(x, y)e_3$. According to the 3D Laplace equation restricted to the open half-space $z > 0$ and the boundary condition of the second kind, we can obtain the monogenic signal as follows:

$$f_M(x, y, z) = p_2 * f(x, y) + p_2 * h_x * f(x, y) + p_2 * h_y * f(x, y), \quad (2.7)$$

where $p_2 = z/2\pi(x^2 + y^2 + z^2)^{3/2}$ is a 2D Poisson kernel and $h_R = (h_x, h_y) = (xe_1/2\pi(x^2 + y^2)^{3/2}, ye_2/2\pi(x^2 + y^2)^{3/2})$ is the Riesz kernel, extension on 2D of the Hilbert kernel, z is equal to the scale space parameter s . Now let $f_M(x, y, s) = f_p(x, y, s) + f_x(x, y, s) + f_y(x, y, s)$ and $f_q(x, y, s) = \begin{bmatrix} f_x \\ f_y \end{bmatrix}(x, y, s)$. In case of intrinsic dimension one signals (The intrinsic dimension expresses the number of degrees of freedom necessary to describe local structure. Constant signals without any structure are of intrinsic dimension zero (i0D), arbitrary orientated straight lines and edges are of intrinsic dimension one (i1D), and all other possible patterns such as corners and junctions are of intrinsic dimension two (i2D). In general i2D signals can only be modeled by an infinite number of superimposed i1D signals.) (i.e., $n = 1$ in (2.1)), we can obtain:

$$(f_p, f_x, f_y) = a(\cos \phi, \sin \phi \cos \theta, \sin \phi \sin \theta), \quad (2.8)$$

where $a(x, y, s)$, $\theta(x, y, s)$, and $\phi(x, y, s)$ represent the local amplitude, local orientation, and instantaneous phase, respectively:

$$\begin{aligned} a(x, y, s) &= \sqrt{f_p^2(x, y, s) + f_x^2(x, y, s) + f_y^2(x, y, s)}, \\ \theta(x, y, s) &= \arctan\left(\frac{f_y(x, y, s)}{f_x(x, y, s)}\right), \\ \phi(x, y, s) &= \arctan\left(\frac{\sqrt{f_x^2(x, y, s) + f_y^2(x, y, s)}}{f_p(x, y, s)}\right). \end{aligned} \quad (2.9)$$

However, we do not know the correct signal of the phase since it depends on the directional sense of $\theta(x, y, s)$. The best possible solution is to project it onto $(\cos \theta, \sin \theta)$ as

$$\hat{\phi}(x, y, s) = \frac{f_q(x, y, s)}{|f_q(x, y, s)|} \arg(f_p(x, y, s) + i|f_q(x, y, s)|). \quad (2.10)$$

2.4. Local Color Phase of the Color Monogenic Signal

In 2009, Demarcq et al. constructed a scale-space signal for color images seen as vectors in $\mathbb{R}_{5,0}$ [4]. Let $s : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ be a real-valued signal and $f : \mathbb{R}^2 \rightarrow \mathbb{R}_{5,0}$ a vector-valued signal, and $\{e_i\}$ ($i = 1, 2, 3, 4, 5$) is the orthonormal basis of \mathbb{R}^5 such that $f(x_1, x_2) = f_3(x_1, x_2)e_3 + f_4(x_1, x_2)e_4 + f_5(x_1, x_2)e_5$. Then a color image is decomposed in the RGB space represented

as the subspace spanned by $\{e_3, e_4, e_5\}$. Now we need to find a function which is monogenic and the e_3 -, e_4 - and e_5 -component, of which are the components r , g , and b , respectively.

$$\begin{aligned}\Delta u &= \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial x_3^2} + \frac{\partial^2 u}{\partial x_4^2} + \frac{\partial^2 u}{\partial x_5^2} = 0, \\ e_3 \frac{\partial u}{\partial x_3} + e_4 \frac{\partial u}{\partial x_4} + e_5 \frac{\partial u}{\partial x_5} &= f(x)\end{aligned}\quad (2.11)$$

with $D = \sum_{i=1}^5 e_i(\partial/\partial x_i)$ and $\Delta = D^2$. A scale-space signal which has independent scales in each component $(f_3, f_4, f_5) = (r, g, b)$ can be defined by splitting the problem into three boundary value problems in $\mathbb{R}_{5,0}$ as follows:

$$\begin{aligned}\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial x_i^2} &= 0, \quad \text{if } x_i > 0, \quad i = 3, 4, 5, \\ e_i \frac{\partial u}{\partial x_i} &= f_i(x_1, x_2) e_i, \quad \text{if } x_i = 0, \quad i = 3, 4, 5.\end{aligned}\quad (2.12)$$

Each solution of the system in (2.12) leads to monogenic functions S_1, S_2, S_3 : they satisfy the Dirac equation in each subspace $E_i = \text{span}\{e_1, e_2, e_i\}$ ($i = 3, 4, 5$) and consequently the Dirac equation in $\mathbb{R}_{5,0}$ ($DS_i = 0$). Let $f_C = S_1 + S_2 + S_3$, then f_C is still monogenic in $\mathbb{R}_{5,0}$ (i.e., $Df_C = 0$) and satisfies the boundary conditions. So the scale-space color monogenic signal can be obtained by using the Dirac operator and the Laplace equation as follows:

$$f_c = \sum_{i=1}^5 A_i e_i = \sum_{i=3}^5 p_2^i * h_x * f_i + \sum_{i=3}^5 p_2^i * h_y * f_i + \sum_{i=3}^5 p_2^i * f_i e_i, \quad (2.13)$$

where $p_2^i = x_i / (2\pi(x_1^2 + x_2^2 + x_i^2)^{3/2})$ ($i = 3, 4, 5$) is a 2D Poisson kernel and $h_R = (h_x, h_y)$ is the Riesz kernel. As for the analytic or monogenic signal, a color image f can be represented in terms of local amplitude and local phase. Now our proposal is to use the geometric product in order to compare two vectors in $\mathbb{R}_{5,0}$.

In the Clifford algebra of the Euclidean vector space \mathbb{R}^n , the product of two vectors a and b , embedded in $\mathbb{R}_{n,0}$, is given by [3, 4]

$$ab = a \cdot b + a \wedge b, \quad (2.14)$$

where $a \cdot b$ is the inner product and $a \wedge b$, the wedge product of a and b , is a bivector. This product is usually called the geometric product of a and b . The inner product is symmetric, and the wedge product is skew symmetric. If $V = ue_1 + ve_2 + ae_3 + be_4 + ce_5 \in R_{5,0}$ is

a chosen vector containing structure information (u, v) and color information (r, g, b) , then the geometric product of f_C and V can be given by

$$\begin{aligned} f_C V &= f_C \cdot V + f_C \wedge V = \langle f_C V \rangle_0 + \langle f_C V \rangle_2 \\ &= \langle f_C V \rangle_0 + \frac{\langle f_C V \rangle_2}{|\langle f_C V \rangle_2|} |\langle f_C V \rangle_2|, \end{aligned} \quad (2.15)$$

where $\langle \cdot \rangle_0$ denotes the scalar part, $\langle \cdot \rangle_2$ the bivector part, and $|\cdot|$ the magnitude of the bivector part [6]. According to Clifford algebras, the geometric product reveals the relationship between bivectors and complex numbers [3]. This means that we can form the equivalent of a complex number, $f_C V = \langle f_C V \rangle_0 + i|\langle f_C V \rangle_2|$, by combining a scalar and a unit bivector. The local color phase can be computed as follows:

$$\varphi_c = \arg(f_C V) = \arctan\left(\frac{|\langle f_C V \rangle_2|}{\langle f_C V \rangle_0}\right). \quad (2.16)$$

This phase describes the angular distance between f_C and a given vector V in $R_{5,0}$, that is, it gives a correlation measure between a pixel fitted with color and structure information and a vector containing chosen color and structure.

3. Local Multi-Model Monogenic Image Feature Descriptors

We make use of a visual representation [11] which is motivated by feature processing in the human visual system and define new local multi-model monogenic image descriptors which give an explicit and condensed representation of the local image signal as follows:

$$\varepsilon = (X, \hat{\phi}, \varphi_c, C). \quad (3.1)$$

In fact, this representation performs a considerable condensation of information in a local image patch of $n \times n$ ($n \in \mathbb{R}^+, n > 1$) pixels. The symbol $X = (x, y)$ represents central coordinates of the local image patch, $\hat{\phi}$ is the instantaneous phase of the gray monogenic signal, φ_c is the local color phase, and $C = (r, g, b)$ is the color values in RGB color space.

Based on the local multi-model image feature descriptors, a stereo similarity function between two primitives is the weighted sum of squared differences of instantaneous phase, local color phase, and color vector for a pair of stereo images in the local patch:

$$\begin{aligned} e(\varepsilon_l, \varepsilon_r) &= \alpha d_{\hat{\phi}}(\hat{\phi}_l, \hat{\phi}_r) + \beta d_{\varphi}(\varphi_l, \varphi_r) + \gamma d_C(C_l, C_r) \\ &= \alpha |\hat{\phi}_l - \hat{\phi}_r| + \beta |\varphi_{cl} - \varphi_{cr}| + \gamma \sqrt{\sum_{i \in \{r, g, b\}} (c_i^l - c_i^r)^2}, \end{aligned} \quad (3.2)$$

where l, r represent the left and right images, respectively, $d_{\hat{\phi}} \in [0, \pi)$ is the distance measurement of instantaneous phase $\hat{\phi} \in [-\pi, \pi)$, $d_{\varphi} \in [0, \pi]$ is the distance measurement

of local color phase $\varphi_c \in [0, \pi]$, and $d_c \in [0, \sqrt{3}]$ is the distance measurement of color vector with $C \in [0, 1] \times [0, 1] \times [0, 1]$ and $c_R + c_G + c_B = 1$ in RGB color space. The symbols α , β , γ are weighted coefficients with $\alpha, \beta \in [0, 1]$ and $\gamma \in [0, 0.5]$. In order to achieve a better coherence with the real scene, we use an adaptive support-weight technique in the local patch [12]. The support weight for each pixel in the window is calculated based on the Gestalt Principles, which state that the grouping of pixels should be based on spatial proximity and chromatic similarity. The original formula proposed is given as follows [13]:

$$w(x, y, m, n) = \exp\left(-\left(\frac{\Delta c_{x,y,m,n}}{\gamma_c} + \frac{\Delta q_{x,y,m,n}}{\gamma_g}\right)\right), \quad (3.3)$$

$$E(x_l, y_l, x_r, y_r) = \frac{\sum_{m,n \in [-p,p]} w_l(x_l, y_l, m, n) \cdot w_r(x_r, y_r, m, n) \cdot e(\varepsilon_l, \varepsilon_r)}{\sum_{m,n \in [-p,p]} w(x, y, m, n) \cdot w_r(x_r, y_r, m, n)}, \quad (3.4)$$

where (x_l, y_l) and (x_r, y_r) are the pixel of interest in left and right images, (m, n) is the pixel offset within the local window, $w(x, y, m, n)$ represents the weight of neighboring pixel $(x + m, y + n)$, $\Delta c_{x,y,m,n}$ is the color difference between pixel (x, y) and $(x + m, y + n)$, $\Delta q_{x,y,m,n}$ is the Euclidean distance between pixel (x, y) and $(x + m, y + n)$, γ_c and γ_g are user defined parameters and, $E(x_l, y_l, x_r, y_r)$ is the aggregated cost between pixel (x_l, y_l) in the left image and (x_r, y_r) in the right image.

In the proposed feature descriptors, there are several merits. Firstly, the instantaneous phase contains local orientation (or geometric information) and information about contrast transition so that it describes an intrinsically one-dimensional structure in a grey level image, that is, an image structure that is dominated by one orientation. Examples of different contrast transitions are a dark/bright (bright/dark) edge or a bright (dark) line on dark (bright) background. Of course, there is a continuum between these different grey level structures. The instantaneous phase as an additional feature allows us to take this information into account (as one parameter in addition to orientation) in a compact way [11]. Secondly, local color phase describes the angular distance between color monogenic signal f_c of color image and a given color vector V in $R_{5,0}$ and gives a correlation measure between a pixel fitted with color and structure information and a vector containing chosen color and structure [14]. Finally, the color vector indicates the mean color structure of local image because color is also an important cue to improve stereo matching. Because the stereo similarity function with the local multi-model feature descriptors contains local geometric, structure, and color information, it is much more robust against noise and brightness change than others in feature matching and 3D reconstruction.

4. Experimentation Results

Once the similarity function is given, minimization process can be performed to find the optimal disparity. In order to reduce noise sensitivity and simultaneously achieve higher efficiency, the multiscale space and a winner-take-all technique are employed to optimize the disparity map for stereo matching. We test the performance of the proposed local multi-model feature descriptors and similarity function on the synthetic and natural stereo images. On the first step, we estimate and compare the disparity maps of a pair of synthetic images using each of three distance measurement and the proposed stereo similarity function. On the

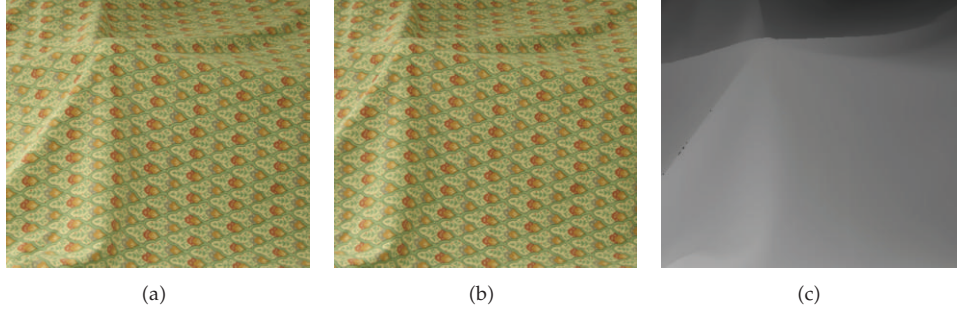


Figure 1: Stereo pair of Cloth1. (a) Left image, (b) right image, and (c) ground disparity map.

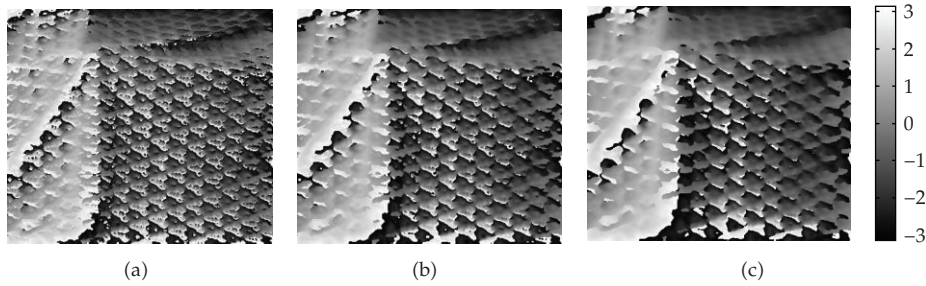


Figure 2: The instantaneous phase maps. (a) First scale, (b) second scale, and (c) third scale.

second step, we reconstruct 3D shape and appearance of natural object and scene combining the proposed stereo similarity function and multiview stereo technique [15].

4.1. Disparity Estimation Experiment

In the first experiment, we chose a pair of color images (cloth1) from the website [16]; the left image, right image, and ground disparity map of the stereo pair are showed in Figures 1(a)–1(c), respectively.

Firstly, the left and right color images are converted to gray level image. The gray monogenic signals with $s = 3$ scales are computed, and the instantaneous phases are estimated for both gray images in Poisson scale space. Figure 2 show three instantaneous phase maps of the left image. At the same time, the scale-space color monogenic signals with $s = 3$ scales are computed, and a chosen reference vector $V = (e_3 + e_4 + e_5)/\sqrt{3}$ with local geometric structure $(u, v) = (0, 0)$ and a unit vector $(a, b, c) = (1, 1, 1)/\sqrt{3}$ in RGB color space are used to estimate the local color phase for both images [14]. Figure 3 also shows three color phase maps of the left image.

Secondly, the proposed stereo similarity function based on local multi-model monogenic image descriptors is employed to compute the weighted sum of squared differences of instantaneous phase, local color phase, and color vector. The scale-space color monogenic framework [14] and a local optimization technique [17] are employed to optimize the disparity map according to the adaptive support-weight cost aggregation as in (3.4). In the scale space, the disparity field at the coarser scale is used as starting guesses for estimation

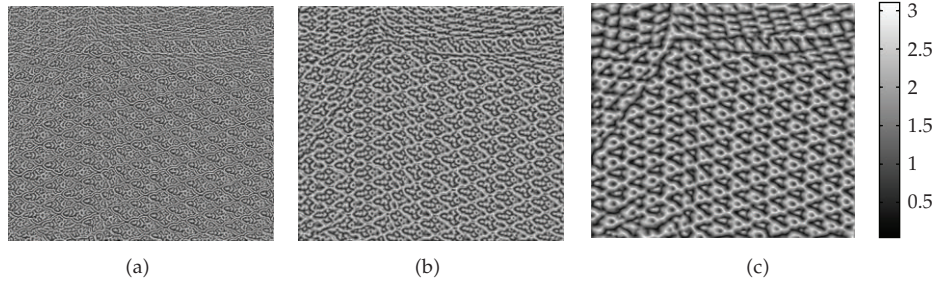


Figure 3: Local-color-phase maps. (a) First scale, (b) second scale, and (c) third scale.

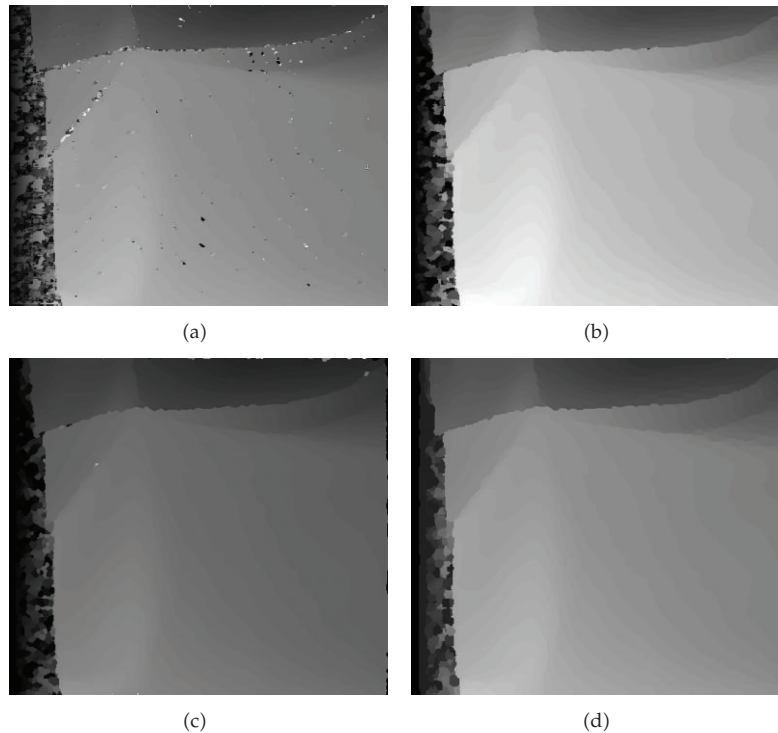


Figure 4: Experiment results for disparity estimation with the different weighted coefficients. (a) $\alpha = 0$, $\beta = 0$, $\gamma = 0.5$; (b) $\alpha = 1$, $\beta = 0$, $\gamma = 0$; (c) $\alpha = 0$, $\beta = 1$, $\gamma = 0$; (d) $\alpha = 1$, $\beta = 1$, $\gamma = 0.5$.

at the next finer scale, then a subpixel disparity can be obtained. For example, we estimate the disparity map at the third scale using the cost function (3.2) and (3.4). The disparities at the third scale directly subtending the second scale under scrutiny are all used as candidate starting guesses. The candidate leading to best match is accepted for the regularization step. Estimation proceeds in this fashion, decrementing scale from coarser to finer, with matching following by regularization, until the finest level of detail is reached. Figure 4 shows the dense disparity map based on the proposed algorithm. Figures 4(a)–4(d) show four estimated disparity maps with the different weighted coefficients (α , β , γ).

Thirdly, a comparison is performed to further validate the claims about the performance of the proposed algorithm. This comparison is performed between the proposed

Table 1: The comparison of the estimated disparity with the different algorithms.

Algorithms	The selected algorithms						The proposed algorithms			
	SAD	SSD	GC	CWT	DWT	QWT	$\alpha = 0$	$\alpha = 1$	$\alpha = 0$	$\alpha = 1$
							$\beta = 0$	$\beta = 0$	$\beta = 1$	$\beta = 1$
							$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0$	$\gamma = 0.5$
RMSE	4.23	3.78	2.25	2.69	3.17	2.53	3.24	2.61	1.98	1.82
PBD	0.37	0.31	0.15	0.19	0.29	0.18	0.30	0.20	0.13	0.09

algorithm and a number of selected algorithms from the literature, which are Sum of Absolute Differences (SAD) [18], Sum of Squared Differences (SSD) [18], Graph Cuts (GC) [18], Discrete Wavelet (DWT) [19], Complex Discrete Wavelet (CWT) [19], and Quaternion Wavelet (QWT) [20]. To create a better understanding of the comparison, statistic root mean square error (RMSE) and percent-age of bad disparities (PBD) are calculated as [19]

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i,j} |d_E(i,j) - d_G(i,j)|^2}, \quad (4.1)$$

$$\text{PBD} = \frac{1}{N} \sum_{i,j} |d_E(i,j) - d_G(i,j)| > \zeta,$$

where d_E and d_G are the estimated and ground truth disparity maps and N is the total number of pixels in an image, whereas ζ represents the disparity error tolerance. The statistics RMSE and PBD related to all of the above algorithms is presented in Table 1. As it can be seen in Table 1, the proposed algorithms with the weighted coefficients $\alpha = 1, \beta = 1, \gamma = 0.5$ obtain better results (RMSE = 1.82 and PDB = 0.09) than others.

4.2. The Robustness against Noise and Brightness

Now we aim to investigate the robustness of the proposed approach against noise and brightness change. In a first step, Gaussian noises with $\sigma = 0.05, 0.10, 0.15, 0.20, 0.25, 0.30$ are added to both color images, respectively, and the disparity maps are estimated using the proposed method with four different weighted coefficients. The statistic root mean square error (RMSE) and percentage of bad disparities (PBD) for each disparity map with noise also are calculated using (4.1). In a second step, the brightness values with $I = -30, -20, -10, 10, 20, 30$ are added to the right image. We also compute the disparity maps, RMSE and PBD, with four different weighted coefficients. The experiment results in Table 2 show that the proposed approach with $\alpha = 1, \beta = 1, \gamma = 0.5$ are unsusceptible to noise and brightness change.

4.3. 3D Reconstruction Experimentation

In the last experiment, the proposed approach is used to reconstruct a set of oriented points for 3D natural scenes by multiview stereopsis. We firstly captured stereo images with the size of 1024×768 around 3D object and scene (a static color paper cup and a deforming color texture paper) under the condition of natural light by two color cameras (AVT Guppy F146C).

Table 2: Experiment results for testing robustness against noise and brightness change.

Noise/brightness change	$\alpha = 0, \beta = 0, \gamma = 0.5$		$\alpha = 1, \beta = 0, \gamma = 0$		$\alpha = 0, \beta = 1, \gamma = 0$		$\alpha = 1, \beta = 1, \gamma = 0.5$	
	RMSE	PBD	RMSE	PBD	RMSE	PBD	RMSE	PBD
None	3.24	0.30	2.61	0.20	1.98	0.13	1.82	0.09
$\sigma = 0.05$	3.46	0.34	2.73	0.22	1.99	0.14	1.89	0.10
$\sigma = 0.10$	3.73	0.37	2.89	0.24	2.13	0.15	1.96	0.11
$\sigma = 0.15$	3.98	0.41	2.99	0.26	2.18	0.16	2.04	0.12
$\sigma = 0.20$	4.41	0.46	3.17	0.28	2.29	0.18	2.12	0.13
$\sigma = 0.25$	5.09	0.51	3.29	0.31	2.41	0.20	2.19	0.15
$\sigma = 0.30$	6.12	0.59	3.50	0.35	2.53	0.22	2.26	0.18
$I = -30$	6.47	0.62	2.86	0.28	2.30	0.19	2.07	0.14
$I = -20$	5.35	0.54	2.67	0.25	2.12	0.16	1.96	0.12
$I = -10$	4.05	0.41	2.43	0.23	1.97	0.14	1.89	0.10
$I = 10$	4.02	0.39	2.39	0.22	1.96	0.13	1.86	0.10
$I = 20$	5.10	0.49	2.46	0.24	2.14	0.15	1.94	0.11
$I = 30$	6.59	0.61	2.72	0.27	2.27	0.18	2.09	0.14

We also calibrate the intrinsic and external parameters for all of cameras and images and estimated the corresponding epipolar lines among images. Feature points for all images are detected by using Speeded-Up Robust Features (SURF) with blob response threshold of 1000 [21] and are matched by using the corresponding epipolar constraints and the proposed similarity function with $\alpha = 1, \beta = 1, \gamma = 0.5$ in (3.2) and (3.4). According to these init matched feature points, we reconstruct a set of 3D sparse oriented points for object and scene similar to multiview stereopsis in [15]. Then we expand and filter it to a set of robust 3D dense oriented points with each image cell of 2×2 pixels and reconstruct. We reconstruct three-dimensional surface of the object. For the static color paper cup, we capture 16 images and reconstruct the 3D dense oriented points and shape. In Figure 5, there are the captured images, the 3D dense oriented points, and three-dimensional surface at the 1st, 7th, and 13th view, respectively. For the deforming color texture paper, we capture the stereo sequence using two cameras at the rate of 7 frames each second and reconstruct the time-varying 3D shape and motion. Figure 6 shows the captured image of one camera and the time-varying 3D shape at the 0th, 20th, 40th, 80th, and 100th frames.

To further validate the performance of the proposed method, a new set of 3D dense patches for the static color paper cup also is reconstructed by the SSD-based photometric discrepancy function in [15]. And, for both algorithms, we calculated the mean feature points (MFP) and percentage of bad disparities (PBD) of each image, the number of 3D oriented points at the init matching (MAT_3D), expanding (EXP_3D) and filtering (FIL_3D) stages, and the percentage of image cells not to be reconstructed (PNR). Statistical results are shown in Table 3.

5. Conclusions

In the paper, we propose a stereo similarity function based on local multi-model monogenic image feature descriptors to solve stereo-matching problem. Local multi-model monogenic

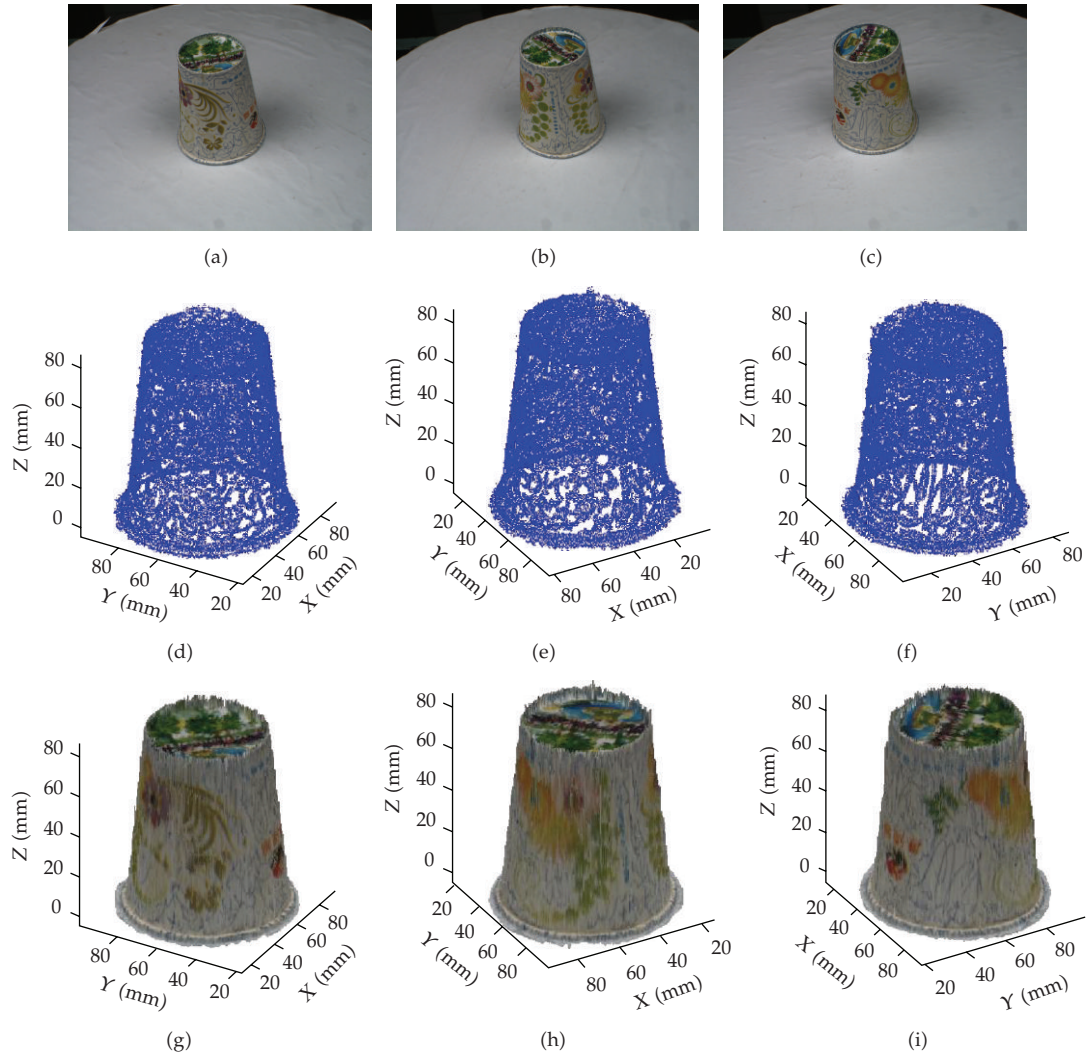


Figure 5: The 3D dense oriented points reconstruction for the color paper cup. (a)–(c) The captured images, (d)–(f) the 3D dense oriented points, and (g)–(i) the three-dimensional surface at the 1st, 7th, and 13th view.

Table 3: Experiment results for matching features and reconstructed 3D oriented points.

Algorithm	MFP	PBD	MAT_3D	EXP_3D	FIL_3D	PNR
The proposed method	1632	0.18	4699	102417	99336	7.26%
SSD-based	1632	0.38	3578	8796	8169	16.32%

image features include local orientation and instantaneous phase of the gray monogenic signal, local color phase of the color monogenic signal, and local mean colors in the multiscale color monogenic signal framework. The gray monogenic signal, which is the extension of analytic signal to gray level image using Dirac operator and Laplace equation, consists of local amplitude, local orientation, and instantaneous phase of 2D image signal. The color

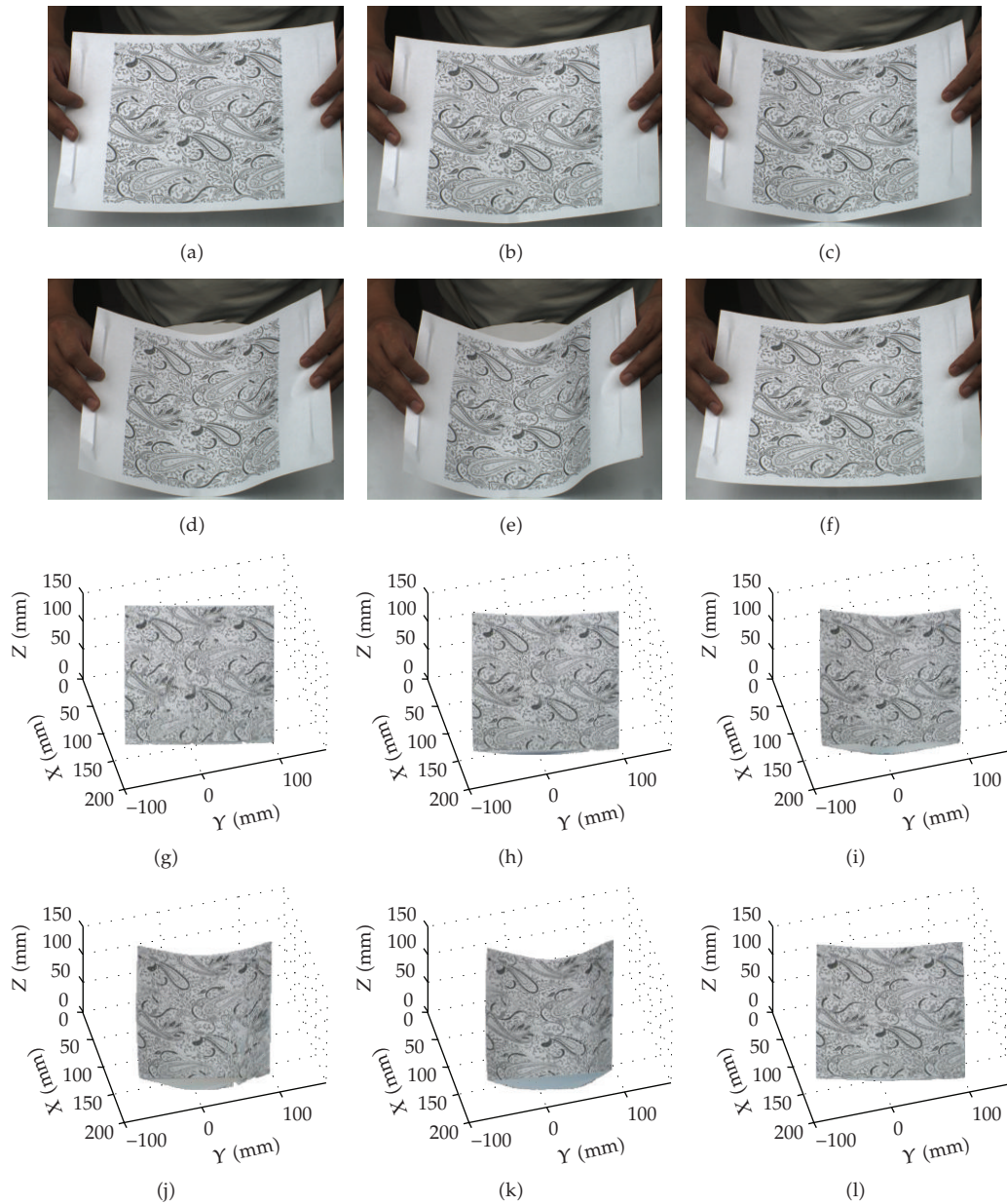


Figure 6: The time-varying shape and motion estimation for the color texture paper. (a)–(f) The captured images and (g)–(l) the three-dimensional surface at the 0th, 20th, 40th, 60th, 80th, and 100th frames.

monogenic signal is the extension of monogenic signal to color image based on Clifford algebras. The local color phase can be estimated by computing geometric product between the color monogenic signal and a unit reference vector in RGB color space. Experiment results on the synthetic and natural stereo images show the performance of the proposed approach. But there is a shortcoming that the proposed method needs much more run time and storage space. So we will be devoted to improving the efficiency and evaluating the proposed method in the future work.

Acknowledgments

This work was supported by National High Technology Research and Development Program (2008AA04Z121), National Natural Science Foundation of China (50975228), and the Technical Innovation Research Program (2010CXY1007) of Xi'an, Shaanxi.

References

- [1] J. Li, H. Zhao, X. Zhou, and C. Shi, "Robust stereo image matching using a two-dimensional monogenic wavelet transform," *Optics Letters*, vol. 34, no. 22, pp. 3514–3516, 2009.
- [2] J. Li, H. Zhao, Q. Fu, and K. Jiang, "Space-time stereo analysis combining local structure and modulation features in the monogenic wavelet domain," *Optics Letters*, vol. 35, no. 7, pp. 1049–1051, 2010.
- [3] J. Vince, *Geometric Algebra for Computer Graphics*, Springer, London, UK, 2008.
- [4] G. Demarcq, L. Mascarilla, and P. Courtellemont, "The color monogenic signal: A new framework for color image processing. Application to color optical flow," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 481–484, November 2009.
- [5] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3136–3144, 2001.
- [6] L. Wietzke and G. Sommer, "The signal multi-vector," *Journal of Mathematical Imaging and Vision*, vol. 37, pp. 132–150, 2010.
- [7] D. Gabor, "Theory of communication," *Journal IEE (London)*, vol. 93, no. 26, pp. 429–457, 1946.
- [8] S. L. Hahn, *Hilbert Transforms in Signal Processing*, Artech-House, Boston, Mass, USA, 1996.
- [9] C. Zetzsche and E. Barth, "Fundamental limits of linear filters in the visual processing of two-dimensional signals," *Vision Research*, vol. 30, no. 7, pp. 1111–1117, 1990.
- [10] M. Felsberg, "Low-level image processing with the structure multivector," Tech. Rep. 2016, Department of Computer Science, Kiel University, Kiel, Germany, 2002.
- [11] N. Kruger and M. Felsberg, "An explicit and compact coding of geometric and structural image information applied to stereo processing," *Pattern Recognition Letters*, vol. 25, no. 8, pp. 849–863, 2004.
- [12] K. J. Yoon and I. N. S. Kweon, "Adaptive support-weight approach for correspondence search," *PAMI*, vol. 28, no. 4, pp. 650–656, 2005.
- [13] Y. Zhang, M. Gong, and Y. H. Yang, "Real-time multi-view stereo algorithm using adaptive-weight parzen window and local winner-take-all optimization," in *Proceedings of the 5th Canadian Conference on Computer and Robot Vision (CRV '08)*, pp. 113–120, May 2008.
- [14] J. Li, H. Zhao, K. Jiang, X. Zhou, and X. Tong, "Multiscale stereo analysis based on local-color-phase congruency in the color monogenic signal framework," *Optics Letters*, vol. 35, no. 13, pp. 2272–2274, 2010.
- [15] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, vol. 1–8, pp. 2118–2125, June 2007.
- [16] <http://vision.middlebury.edu/stereo/data/scenes2006/FullSize/Cloth1/>.
- [17] L. Wang, M. Gong, M. Gong, and R. Yang, "How far can we go with local optimization in real-time stereo matching - A performance study on different cost aggregation approaches," in *Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT '06)*, pp. 129–136, June 2006.
- [18] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1–3, pp. 7–42, 2002.
- [19] A. Bhatti and S. Nahavandi, "Stereo image matching using wavelet scale-space representation," in *Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation (CGIV '06)*, pp. 267–272, July 2006.
- [20] Y. I. Xu, X. Yang, P. Zhang, L. I. Song, and L. Traversoni, "Cooperative stereo matching using quaternion wavelets and top-down segmentation," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '07)*, pp. 1954–1957, July 2007.
- [21] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

