A multiple view polarimetric camera

Article (Accepted Version)

A Multiple View Polarimetric Camera

PHILIP BIRCH[1,*], RUPERT YOUNG[1], AND CHRIS CHATWIN[1]

[1] Department of Engineering and Design, University of Sussex, Falmer, BN1 9QT, United Kingdom
[*] Corresponding author: p.m.birch@sussex.ac.uk

Compiled July 3, 2018

A multiple view polarimetric camera is developed. The system uses four separate action cameras and software is employed to map the images onto each other in order to generate the Stokes vectors, the degree of linear polarisation and angle images. To ensure robustness, an automated calibration system has been developed that ensures the pixels are correctly mapped. Video frame synchronisation is also developed.

OCIS codes: (110.5405) Polarimetric imaging, (120.5410) Polarimetry

http://dx.doi.org/10.1364/ao.XX.XXXXXX

INTRODUCTION

The ability to image a scene and capture the reflected polarisation state has numerous applications: for target detection [1–5]; to material identification[6]; to imaging through fog[7–10] and rain [11]. There are a large number of remote imaging applications [12], and it is often measured in the infra-red region [13]. There have been a number of designs of polarimetric cameras in the recent years. Commercial cameras are available but the different trade-offs in design means there is no single solution to polarimetric imaging and the area is still subject to research. The requirement to take multiple images to capture the polarisation state makes the process slightly more involved than capturing an intensity image. Typically four measurements are made to capture the linear polarisation states, plus additional measurements to capture the circular states. The need for multiple images gives raise to the large number of camera designs.

The overall type of camera to be designed is dependent on the application. To measure all the Stokes vectors over a wide range of wavelengths is desirable but it requires spectropolarimetric imaging [14–16]. By using wave-plates and bandpass filters the polarimetric image can be measured at a set of wavelengths using imaging camera technology by taking a number of images with a known polariser or wave-plate in front of the camera. By only measuring the linear polarisation states, the camera design can be further simplified, but with the loss of information about the circular polarisation state.

To capture the multiple images required to image polarisation states there have been many different designs of cameras. Since multiple measurements are required they can be broken down in two major types: temporal multiplexing cameras (or division of time) [17–19] and spatially multiplexing (or division of wave-front) [9, 20]. The temporal multiplex devices typically have a single camera sensor and record N images with a different polarisation modulation imposed on each. The states are often modulated using liquid crystal birefringence or a rotation of some polarisation element.

Division of wave-front cameras typically work by splitting the image into N optical paths and imaging onto these onto different sensors that measure differing polarimetric states. The major drawback with the temporal multiplexing devices is that the frame rate becomes limited by the requirement to make N measurements per result. With wave-front splitting devices the frame rate can be high but there is a potential loss of light due to the splitting of the wave-front and there is the additional cost of multiple sensors to be housed in one camera.

A variation of this design is the 'division by focal plane' (DoFP) were a single sensor is used. Multiple camera pixel sized filters are combined together into a macro-pixel to perform a measurement. This often requires polarising elements to be placed directly on the sensor [21].

Many of the designs of polarimetric cameras suffer from complicated optics which increases the bulk of the camera, expense and robustness. In addition, the field of view can become restricted due to the additional components in the optical path. Some work, such as that in Reference 22 have attempted to ameliorate this.

Both the division by wave-front and division by time necessitate the division of the incident light field. In both cases the amount of light is decreased by N for N measurements. With the additional requirement for linear polarisation filters, this has a negative impact on the overall systems signal to noise ratio.

The required overall accuracy of the measurements made is dependent on the application. Metrology measurement applications in which material are studied requires a high degree of accuracy. For robotics application in which the device might be used for water detection or object detection the constraints maybe relaxed slightly since the camera is only part of larger processing chain and a computer vision process. From a drone

looking downwards water will have a higher degree of polarisation, since it is reflecting the sky, compared to surrounding vegetation which will be large depolarised. In this case a measurement accuracy of 5% or even higher may well suffice.

In this paper we are taking a different approach to other designs. We develop a light weight low cost, linear state polarimetric camera. The design is for robotic drone navigation applications but not restricted to robots. Due to the application, the device must be robust, light weight and low cost. The accuracy of device does not need to be high since it is intended to be used as a navigational aid such as for detecting water. Since the robot will be moving, the temporal multiplexing device has been ruled out. To maintain a low cost and wide field of view we have developed a polarimetric system that uses a set of independent cameras to make separate measurements, and the results are calculated in post-processing. The cameras are typically sold as sports action cameras or vehicle dashboard cameras. Each camera costs about US$100 and only requires the addition of a linear polariser, making this a low cost system to buy. The construction is also simpler than other types of camera such as a DoFP camera. The camera can be built from scratch in few hours with no specialised equipment. Since multiple cameras are used, we do not decrease the amount of light to each image by N.

In the next section we will discuss the background and cover how Stokes parameters are measured. Then in the Camera Design section we will cover the camera's design and how the images are mapped onto each other. Analysis of camera noise and the problem of having independent cameras is discussed in Noise Errors. How the videos are synchronised is covered in the Video Synchronisation section and finally some example results are shown in the Results.

BACKGROUND

A polarised light wave, travelling along the z axis can be described as a sum of two orthogonal electromagnetic waves.

$$\mathbf{E}(z,t) = E_x\hat{\mathbf{x}} + E_y\hat{\mathbf{y}} \tag{1}$$

where

$$\begin{aligned} E_x &= a_x\cos\left(\omega\left(t - z/c\right) + \phi_x\right) \\ E_y &= a_y\cos\left(\omega\left(t - z/c\right) + \phi_y\right) \end{aligned} \tag{2}$$

where ω is the angular frequency, $(t - z/c)$ is the time varying component, a_x and a_y are the amplitudes, and ϕ_x and ϕ_y are the absolute phases of the wave components. The variables a_x, a_y, ϕ_x and ϕ_y allow us to define the different linear, elliptical or circular polarisation states. It is convenient to then reformulate these into a Stokes vector $\mathbf{S} = [s_0, s_1, s_2, s_3]^T$ where:

$$\begin{aligned} s_0 &= a_x^2 + a_y^2 \\ s_1 &= a_x^2 - a_y^2 = P\cos 2\alpha \\ s_2 &= 2a_xa_y\cos\phi = P\sin 2\alpha \\ s_3 &= 2a_xa_y\sin\phi \end{aligned} \tag{3}$$

where P is the degree of linear polarisation and α is the angle of polarisation. From this we see s_0 is the total intensity of the beam, s_1 is the difference between the two intensities, and s_2 and s_3 are proportional to the real and imaginary parts of the total wave. These parameters can be experimentally measured by taking N number of intensity measurements with different, but known, polarisation elements placed in front of the sensor. The Stokes vector can then be determined by some fitting algorithm. In the simplest case, we can label the measurements as I_q where q is either: θ, the angle of a linear polariser placed in front of the camera or the left (L) or right (R) hand circular polarisation measurement. The measured Stokes vectors are as follows:

$$\begin{aligned} \hat{s_0} &\propto I_0 + I_{90} \\ \hat{s_1} &\propto I_0 - I_{90} \\ \hat{s_2} &\propto I_{45} - I_{135} \\ \hat{s_3} &\propto I_L - I_R \end{aligned} \tag{4}$$

The measurement of s_3 is more problematic than the others since it is usually performed by introducing a wave-plate which then introduces wavelength dependency issues. For this reason, the parameter is often neglected. If all four elements of the Stokes vector are measured this is known as a 4D camera. By neglecting s_3, it is then known as a 3D camera. If only two measurements are made, only s_0 and s_1 can be calculated and this known as a 2D camera [12].

For the 3D case, the derivation of Equation (4) can be obtained by rearranging Equation (3) to get the i^{th} measurement intensity [23]:

$$I_i = \frac{s_0}{2N}\left[1 + P\cos(2\theta_i - 2\alpha)\right] \tag{5}$$

This can be expressed , with the addition of noise, in a matrix form:

$$\mathbf{I} = \frac{1}{N}\mathbf{W}\mathbf{S} + \mathbf{b} \tag{6}$$

where $\mathbf{b}$ is Gaussian additive noise with zero mean and σ^2 variance and $\mathbf{W}$ is an $N \times 3$ matrix:

$$\mathbf{W} = \frac{1}{2}\begin{bmatrix} 1 & \cos 2\theta_1 & \sin 2\theta_1 \\ & \vdots & \\ 1 & \cos 2\theta_N & \sin 2\theta_N \end{bmatrix} \tag{7}$$

The measured Stokes parameters can then be calculated by finding the pseudo-inverse of $\mathbf{W}$, i.e.:

$$\mathbf{W}^\dagger = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T \tag{8}$$

and

$$\hat{\mathbf{S}} = N\mathbf{W}^\dagger\mathbf{I} \tag{9}$$

The N factor appears in these equations because the total amount of light is reduced for every measurement. This is the case for both wave-front and time division polarimetric cameras.

It can be shown [23, 24] that the optimum angles that minimise the covariance of the estimation is:

$$\theta_i = \pi\frac{i-1}{N} \tag{10}$$

and the covariance is given by

$$\mathbf{\Gamma}^{\hat{\mathbf{S}}} = N^2\mathbf{W}^\dagger\mathbf{\Gamma}^{\mathbf{I}}(\mathbf{W}^\dagger)^T = 4N\sigma^2\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{11}$$

where $\mathbf{\Gamma}^{\mathbf{I}} = \sigma^2\mathcal{I}_N$ is the covariance matrix of the measurements and assumed to be additive noise, with variance σ^2. $\mathcal{I}_N$ is an identity matrix. This means that increasing the number of measurements, actually increases the error. This result can be understood by inspecting Equation (9), as the number of measurement

are made, the amount of light impinging on the sensor is decreased, whilst the noise, **b**, remains constant.

From the measurements of the values in Equation (4), a number of common metrics can be derived. If only two measurements are made with $\hat{I}_0$ and $\hat{I}_{90}$, only $\hat{s}_0$ and $\hat{s}_1$ can be determined. From this we get the polarisation contrast (PC):

$$PC = \frac{s_1}{s_0} \tag{12}$$

The degree of linear polarisation (DOLP) is the fraction of the light attributed to the linearly polarised part and is given by:

$$DOLP = \hat{P} = \frac{\sqrt{s_1^2 + s_2^2}}{s_0} \tag{13}$$

The total degree of polarisation (DOP) requires knowledge of all four Stokes parameters:

$$DOP = \frac{\sqrt{s_1^2 + s_2^2 + s_3^2}}{s_0} \tag{14}$$

The angle of polarisation is the angle of the elliptical major axis to the x axis and is given by:

$$\phi = \frac{1}{2}\arctan\left(s_2/s_1\right) \tag{15}$$

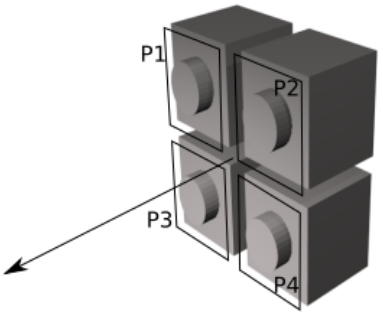


Fig. 1. 4 camera design with the system optical axis indicated. P1, P2, P3 and P4 are linear polarisers that have been placed in front of each camera. Each is at a different angle.

CAMERA DESIGN

We have N cameras where $N \in \{2,3,4\}$, giving either a 2D or 3D Stokes parameter measurement. These cameras are placed as close as possible in the same plane. The four camera layout is shown in Figure 1. The system has an overall chief optical axis as indicated in Figure 1 and four translated and potentially rotated optical axes for each camera.

The cameras are Mobius ActionCams that can record at 1080p (1920×1800) at 60 fps. These cameras are typically used as car dashboard cameras (dash cameras) or as body worn action cameras. The video is compressed using H264 AVC with a variable bit rate. The images, with audio, can be sent directly to a host computer via USB or alternatively stored on an internal SD card. The cameras were set up to record to SD card as soon as they were powered on. The choice of using the USB bus or SD-card enables the cameras to be used in either on-line or off-line mode.

In order to calculate the Stokes vector for each pixel in the image, each camera image must be aligned with pixel perfect accuracy. Unless we are imaging objects at a large distance, this becomes a non-trivial problem and so a projection matrix that maps one pixel to another is used. Since the cameras are not on the systems optical axis we select the most central image, which we call the prime camera, and project the other camera images onto the prime image. The co-ordinates of each pixel in camera m are x_m and y_m and these are expressed in homogeneous image coordinates as $\mathbf{x} = (x_1, x_2, x_3)^T$ where $x = x_1/x_3$ and $y = x_2/x_3$.

$$\acute{\mathbf{x}} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^T & 1 \end{bmatrix} \mathbf{x} = \mathbf{H}\mathbf{x} \tag{16}$$

Here $\mathbf{A}$ is a 2×2 matrix describing the affine transform, $\mathbf{t}$ is 2×1 vector describing the spatial translation shift and $\mathbf{v}$ is a 2×1 vector describing the projective non-linearities.

If p is the prime camera number, we have a set of homography matrices to discover:

$$\acute{\mathbf{x}}_n = \mathbf{H}_{p,n}\mathbf{x}_n \tag{17}$$

where $n \in 1, \ldots N, n \neq p$. Once the set of homography matrices are discovered, we can use interpolation to correct for the projective distortion and produce a set of images $I_n(\acute{x}, \acute{y})$ and the prime image $I_p(x, y)$.

The cameras also have a degree of distortion due to the lens. This is corrected by discovering the intrinsic matrix, $\mathbf{K}$, for each camera in turn before the homography is calculated [25].

Since the cameras have different optical axes, there will be a parallax problem when viewing objects close to the cameras. The object will appear in different positions depending on its depth, like in a stereopsis system. Unless this can be corrected it presents a limitation of the system. If we consider two cameras as simple pinhole models, and one camera shifted in the plane orthonormal to the optical axis. The two camera matrices are:

$$\begin{aligned} \mathbf{P_p} &= \mathbf{K}\left[\mathbf{I}|\mathbf{0}\right] \\ \mathbf{P_n} &= \mathbf{K}\left[\mathbf{I}|\mathbf{t}_n\right] \end{aligned} \tag{18}$$

where

$$\mathbf{K} = \begin{bmatrix} f & & 0 \\ & f & \\ 0 & & 1 \end{bmatrix} \tag{19}$$

where f is the focal length of the lens, $\mathbf{I}$ is an identity matrix, and $\mathbf{t}_n$ is the translation matrix $[-t_x, 0, 0]^T$. The position of an objects' image in each camera can be calculated then from:

$$\begin{aligned} \acute{\mathbf{x}}_n &= \mathbf{P}_n\mathbf{x} \\ \acute{\mathbf{x}}_p &= \mathbf{P}_p\mathbf{x} \end{aligned} \tag{20}$$

If we set the object's position to be halfway between the two optical axes, $x = t_x/2$ and inserting this into Equation (20) we get:

$$z = \frac{f t_x}{2\acute{x}} \tag{21}$$

Setting $\acute{x}$ to the size of one camera pixel gives a measure of $z = z_{\min}$, the minimum distance an object can be located at and not give errors in the polarisation calculation due to parallax.

The cameras used in our experiment use a Aptina ARO330 sensor with a $2.2\mu m$ pixel pitch and $2.1mm$ focal length lens. The resolution of the sensor is 2304×1536. Due to the camera housing the separation of the optical axes is $t_x = 40mm$, giving a $z_{\min} = 19m$. However, by pixel binning to reduce the resolution to 640×480, this is reduced to $z_{\min} = 5.2m$. To work closer than that, either the parallax has to be corrected in software or assumptions made that the imaged object has slowly varying Stokes parameters. This maybe suitable for some applications.

For the case with four cameras, a stricter constraint might be imposed. Since the cameras are arranged in a two dimensional array and fitting the results within $\frac{1}{2}$ a pixel means the maximum error is $\frac{1}{\sqrt{2}}$. This increases the minimum range to $7.5m$.

It should also be noted that since each individual camera is at a slightly different angle the polarisation state of the light will be slightly different for each. However, for this camera at the minimum range of $5.2m$, the angular difference is only 0.44°. If the cameras are operated at a closer range, this will need consideration.

Homography discovery

Homographic mappings between multiple two-dimensional planes such as images is a well developed technique and has many applications such as photograph stitching and stereo camera mapping [26, 27]. The typical methodology is to use SIFT or another key-point matching method to find correspondences between the set of images. An outlier rejection method such as RANSAC [28] is then applied. The good matches are then optimised by minimising the geometric error cost:

$$d_\perp^2 = \sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\hat{\mathbf{x}}, \mathbf{H}\hat{\mathbf{x}}_i)^2 \quad \textbf{(22)}$$

over $\mathbf{H}$ and $\hat{\mathbf{x}}_i$ for the set of matches using the Levenberg-Marquardt algorithm.Here $i = 1, 2, \ldots, n$, where n is the number of 2D coordinates selected by RANSAC. This method has been successfully demonstrated on conventional images. SIFT features will be discovered on the four I_q images (Eqn. 4), but we cannot reliably guarantee correspondence between matches across the four images unless the DOLP is zero across the image. Since the DOLP refers to a single pixel we define a new parameter that describes the whole image, the Ratio of Linear Polarisation (ROLP), as:

$$\text{ROLP} = \frac{1}{M}\sum_i^M DOLP_i \quad \textbf{(23)}$$

For images with ROLP ≈ 0, there should be a large number of SIFT matches since $I_0 \approx I_{45} \approx I_{90} \approx I_{135}$. For images with a higher ROLP, the I_q images will be significantly different in appearance. This presents a potential problem. The calibration could be performed with a known low ROLP scene but this could be potentially difficult to implement.

To discover a good $\mathbf{H}$ it is vital that the sample points are well spread across the image since the error increases as the fourth power with distance [p. 149 26]. With a polarised scene there is a chance that some regions will produce no matches. In this case the minimisation of Equation (4) does not always yield good results due the poor spread of $\mathbf{x}_i$, the presence of noise in the localisation of $\mathbf{x}_i$ and lack of machine precision problems in its calculation. For this reason we take a whole image approach to the bundle adjustment problem.

To start, we introduce another outlier removal process and then a global optimisation is performed that is not reliant on SIFT key-points. Since the calibration is running on a video stream, rather than a still image, we have multiple estimates for $\mathbf{H}$, one per frame. We then perform a cross-correlation on the image edges to ensure a good potential match. Cross-correlation is well suited as a metric since it integrates the whole image rather than relying on key-points. This makes it robust to noise. Since cross-correlation is not rotation, scale or affine invariant it is a suitable metric since if the homography correction is valid, there should be none of these transforms present.

The images are pre-filtered with a Difference of Gaussian (DoG) filter. This has the effect of removing high frequency noise and enhancing the edges of an image. The DoG filtering is only performed as part of the calibration process and is not used to calculate the Stokes vectors. By selecting a specific σ value for the DoG, different sized features can be enhanced if desired [29] to tune the sensitivity of the correlation. A normalised cross-correlation is then performed at pixels u and v to produce a new two dimensional array:

$$\gamma(u,v)_{p,n} = \frac{\sum_{x,y}\left[I_n(x,y) - \bar{I}_n\right]\left[I_p(x-u, y-v) - \bar{I}_p\right]}{\left\{\sum_{x,y}\left[I_n(x,y) - \bar{I}_n\right]^2 \sum_{x,y}\left[I_p(x-u,y-v) - \bar{I}_p\right]^2\right\}^{1/2}} \quad \textbf{(24)}$$

where the bar indicates the mean of the image and $n \in \{1, \ldots N\}, n \neq p$. A larger σ for the DOG filters, will boost the higher frequency components of the image, making Equation (24) more sensitive.

In practice, there are many frames in which the SIFT estimate fails. To detect this we calculate $\gamma(u,v)_{p,n}$ for each $\mathbf{H}_{p,n}$ and then calculate:

$$\Gamma_{p,n} = 1 - \max(\gamma(u,v)_{p,n}) \quad \textbf{(25)}$$

We test if $\Gamma_{p,n}$ is less than a threshold. The candidate homography arrays are then used as a starting point for a quasi-Newton minimisation method [p. 136 30] which minimises Equation (25). We will call this method the DoG filtered cross-correlation minimisation (DXM) algorithm.

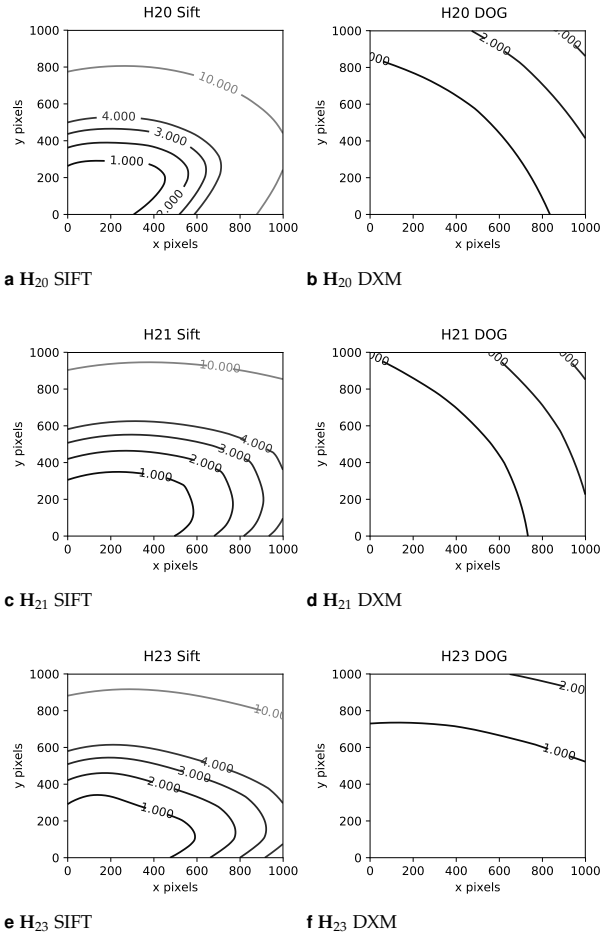
Homography Error

By imaging a static object we can get an estimate of the error in the homography matrix. To test the reliability of the mapping method, the measurements were measured over multiple frames. Due to the RANSAC process the variation is non-Gaussian and in addition there is a correlation between the elements in $\mathbf{H}$. To model this variation, 50 new frames were compared against a manually selected key frame. The variation from the key frame for each pixel was calculated with the algebraic error:

$$d_{alg}^2 = \sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i) \quad \textbf{(26)}$$

and the results are shown in Figures 2 a), c) and e) for each projection using SIFT and RANSAC alone. In order for the polarimeter to give a good result, the d_{alg}^2 pixel error must be less than one. It can be seen from Figures 2 a), c) and e) that this results in an effective resolution of only 250×305 pixels.

To improve on this, the DXM algorithm was applied, taking the SIFT estimated values as the initial starting position. Figures 2 b), d) and f) show the pixel errors. The additional step increases the resolution to about 600×600 pixels.



a $\mathbf{H}_{20}$ SIFT **b** $\mathbf{H}_{20}$ DXM

c $\mathbf{H}_{21}$ SIFT **d** $\mathbf{H}_{21}$ DXM

e $\mathbf{H}_{23}$ SIFT **f** $\mathbf{H}_{23}$ DXM

Fig. 2. $\mathbf{H}_{pn}$, with $p = 2$, errors produced by the SIFT calibration method and DXF method

NOISE ERRORS

There are a number of possible error sources from the camera design. The cameras all have a degree of additive Gaussian white noise (AGWN). It was shown earlier (Equation (11)) that with most camera designs, the impact on the measurement of **S** increases with the number of measurements due to the division of the intensity. In this design this is not the case. However, since the cameras are completely independent, they all have a different amount of auto-gain which will introduce errors into the measurements.

Auto-gain error

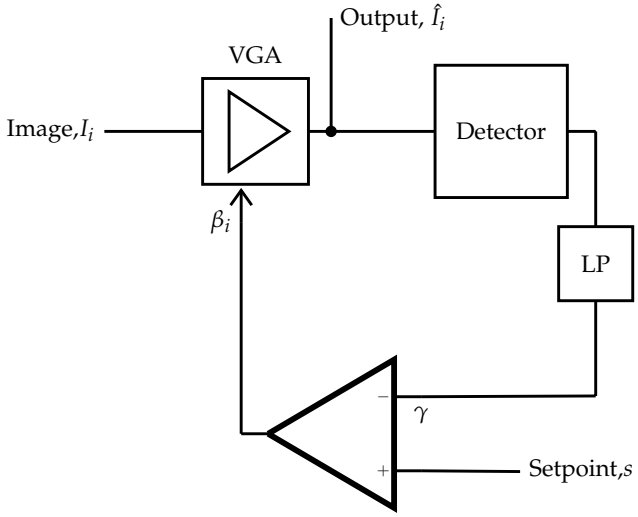


Fig. 3. Auto-gain circuit model: VGA: Varible gain amplifier, LP: low pass filter

Since each camera is a separate device, they are each subject to an independent auto-gain circuit (AGC). For a completely unpolarised image, each I_q image will be identical apart from noise. However, as the degree of polarisation within the image increases, each I_q image will diverge and be subject to a different amount of auto-gain.

Since this auto-gain is unknown, and possibly different for each model of camera, we model this as shown in Figure 3. The raw image data has a gain amplification applied by the variable gain amplifier (VGA). The gain is calculated by taking the low-passed intensity of the voltage (which is equal to the image mean). The gain,β, is then such that $\gamma = s$. In addition, there is clipping which insures the range of the output is between zero and one. A Monte-Carlo simulation is then performed to analyse the model. For a specific value of ROLP a set of DOLP arrays, $\mathbf{P}_i$, can be generated with uniform random numbers. $i \in (1, \ldots 100)$ and:

$$\mathbf{P}_{i,k} \backsim \mathcal{U}(b_i, t_i)_k \tag{27}$$

where b and t are lower and upper bounds of the uniform random number generator, k is the pixel coordinate, and:

$$b_i = \max(0, 2 \times ROLP_i - 1) \tag{28}$$

$$t_i = \begin{cases} 2 \times ROLP_i & \text{if } 2 \times ROLP_i - 1 < 0 \\ 1 & \text{otherwise} \end{cases} \tag{29}$$

From this the Stokes vectors and individual images can be randomly generated such that they give the correct DOLP value

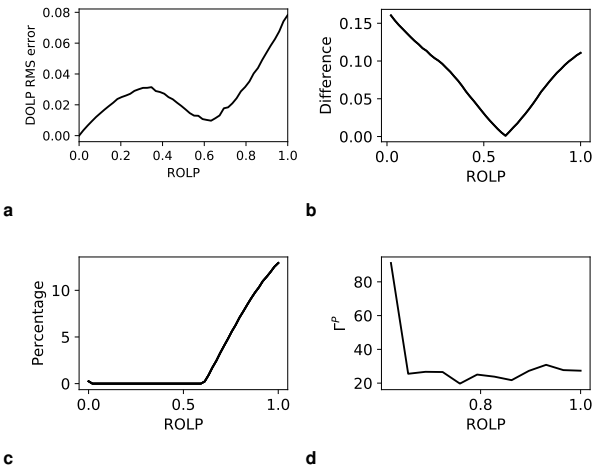


Fig. 4. Simulated errors due to the AGC
(a)DOLP Error; (b)Mean error between simulated image and AGC corrected image; (c)Percentage of the images clipped by the AGC; (d)Simulated mean $\Gamma^{\hat{P}}$ against ROLP

for a pixel. This is achieved by rearranging Equation (13) and generating random numbers with the correct bounds.

By then applying the gain circuit (Figure 3) to each image the new DOLP can be calculated. The RMS error from this and **D** can be calculated. This is repeated 100 times to get an average error for a given ROLP. Figure 4a shows the result of the error. There are a number of factors that explain the shape.

1. For a ROLP=0, each of the four cameras will be identical, and therefore experience equal gain. The error should then be 0.

2. As the ROLP increases, there will be an increasing difference between the images and so the error will therefore increase. This can be seen in Figure 4b which shows the mean difference between the raw simulated image and the AGC corrected images as the ROLP increases. There is a minimum difference, which coincides with DOLP error minima in Figure 4a.

3. The set-point value is $s = 0.5$. As the ROLP tends towards 0.5, the mean of the uncorrected images also tends towards 0.5. This means the AGC will introduce less error, and so the DOLP overall error decreases.

4. Above 0.5, there is an increasing amount of clipping since the pixel values of the images multiplied by the gain are greater than one. This and point 2, cause the error to increase again. The percentage of the clipped pixels as the ROLP increases is shown in Figure 4c. This error contributes to the steeper rise in the DOLP error after 0.6 in Figure 4a.

From Figure 4a, the maximum error due to the AGC is 7.8% with a mean value of 2.7%.

The output intensity for of a pixel can be modelled using Figure 3 as:

$$\hat{I}_i = \beta_i I_i = \frac{s}{\bar{I}_i} I_i \tag{30}$$

where $\bar{I}_i$ is the mean intensity of the entire i^{th} image and largely independent of I_i for a large number of camera pixels. If we assume that each camera is subject to the same amount of AGWN with variance σ^2 and each camera is uncorrelated, the covariance matrix for the set of 4 cameras is:

$$\mathbf{\Gamma}^{\hat{\mathbf{I}}} = \sigma^2 \begin{bmatrix} \beta_1 & 0 & 0 & 0 \\ 0 & \beta_2 & 0 & 0 \\ 0 & 0 & \beta_3 & 0 \\ 0 & 0 & 0 & \beta_4 \end{bmatrix} \tag{31}$$

The covariance of the Stokes vector is then:

$$\Gamma^S = \mathbf{W}^{\dagger} \mathbf{\Gamma}^{\hat{I}} (\mathbf{W}^{\dagger})^T \tag{32}$$

$$= (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{\Gamma}^{\hat{I}} \mathbf{W} \mathbf{W}^{-1} (\mathbf{W}^{-1})^T \tag{33}$$

and for $N = 4$:

$$\mathbf{W} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \tag{34}$$

$$\mathbf{W}^T \mathbf{W} = \frac{1}{4} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{35}$$

so

$$\mathbf{\Gamma}^{\hat{\mathbf{I}}} = \frac{\sigma^2}{4} \begin{bmatrix} \beta_1 + \beta_2 + \beta_3 + \beta_4 & 2\beta_1 - 2\beta_3 & 2\beta_2 - 2\beta_4 \\ 2\beta_1 - 2\beta_3 & 4\beta_1 + 4\beta_3 & 0 \\ 2\beta_2 - 2\beta_4 & 0 & 4\beta_2 + 4\beta_4 \end{bmatrix} \tag{36}$$

If $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$, this simplifies to:

$$\mathbf{\Gamma}^{\hat{\mathbf{I}}} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{37}$$

This is similar to Equation (11), but with the key difference that, because we are not dividing the light field, the noise term does not increase with the number of measurements, N.

From Equation (36) we can see there is an amount of cross-talk between terms due to the separate β variables. However, for images with only a small ROLP, the β terms will be similar in value, so this is potentially low.

The covariance of the DOLP can be estimated by assuming:

$$\mathrm{Var}(y) \approx \nabla f(< x >)^T \, \mathrm{Var}(x) \nabla f(< x >) \tag{38}$$

where $< y > \approx f(< x >)$. Using P to represent the DOLP:

$$\nabla P = \frac{1}{P s_0^2} \left[-P^2 s_0, s_1, s_2 \right]^T \tag{39}$$

$$\Gamma^{\hat{P}} = \frac{\sigma^2}{4 P^4 s_0^4} \Big\{ {s_1}^2 (\beta_1 + \beta_3) + {s_2}^2 (\beta_2 + \beta_4) + P^4 s_0^2 [\beta_1 + \beta_2 + \beta_3 + \beta_4] + 2P^2 s_0 [s_1 (\beta_3 - \beta_1) - s_2 (\beta_2 - \beta_4)] \Big\} \tag{40}$$

If the auto-gain is locked for all cameras, such that $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$, then Equation (40) reduces to:

$$\Gamma^{\hat{P}} = \frac{4\sigma^2}{s_0}\left(2 + P^2\right) \tag{41}$$

The variance in the angle can be approximated by substituting the gradient of Equation (15):

$$\nabla\hat{\alpha} = \frac{1}{2P^2 s_0^2}\left[0, -s_2, s_1\right]^T \tag{42}$$

into Equation (38) and we get:

$$\Gamma^{\hat{\alpha}} = \frac{\sigma^2}{16P^4 s_0^4}\left(s_1^2\left(\beta_2 + \beta_4\right) + s_2^2\left(\beta_1 + \beta_3\right)\right) \tag{43}$$

Due to the approximation made in Equation (38), Equation (43) and Equation (40) are only valid for a reasonably large value of P. For a large value of P the β variables will also be different from each other. Figure 4d show simulated results using the Monte Carlo data generated above to find values for β and the Stokes parameters. For lower values of P the simulation gives unreasonably large values, supporting the assertion that the Equation (38) approximation is only valid for large values of P. From Figure 4d, it can be seen that the $\Gamma^{\hat{P}}$ is actually reasonably constant and the so the AGC has little effect on the overall noise. For lower values of ROLP, the images should be largely unpolarised and therefore similar in intensity. For this case $\beta_1 \approx \beta_2 \approx \beta_3 \approx \beta_4$ and so, from Equation (41), the noise will largely depend on the overall intensity of the image.

The angle measurement is less effected by the differing gains than the DOLP measurements due to the additional terms in Equation (40).

VIDEO SYNCHRONISATION

In off-line mode each camera records onto an SD-card and it is not possible to accurately time stamp the video. The videos are not synchronised and so there is a random frame difference that occurs between each camera. To discover these frame differences the audio track was used. The cross-correlation of the audio was then taken to find the frame differences. First, the audio of the j^{th} video, A_j, is extracted, $j \in [1, \ldots N]$. A window of W audio samples is extracted is used to calculate:

$$O_j = \underset{t}{\operatorname{argmax}}\left\{\int_0^W A_1(t)A_l(t+\tau)d\tau\right\} \tag{44}$$

where $l = 2, \ldots N$. The synchronisation of each video frame was then achieved by skipping $O_j - \min(O_l)$ frames for each corresponding stream.

The accuracy of the above calculation is only $\pm\frac{1}{2}$ time steps. However, the audio is recorded at a considerably faster rate of $44.1kHz$ as compared to the video rate of $25Hz$. This means using this method, the video can be synchronised to $\pm 13\mu s$ or 0.057% of a video frame. Two videos should therefore stay within synchronisation for about $70s$. This could be further improved by two methods. To calculate Equation 44, fast Fourier transforms are used and these could be zero padded to give an arbitrary precision - although in reality this will become limited by machine precision or signal noise. To improve the calculation time and reduced the RAM overheads other methods could be adapted to perform this if needed (reference 31 could be adapted to perform this). Instead, provided there is sufficient audio signal (e.g., music, audio, background noise), W only needs to capture a few seconds. Therefore, the accuracy is perfectly good enough since Equation 44 can be recalculated every few seconds to ensure that the videos remain in synchronisation. This latter method was used in practice.

RESULTS

The device was calibrated by locking the auto-gain, and white balance setting of each of the cameras and illuminating each camera with a known rotatable polarisation state. From intensity fitting on each camera the $\mathbf{W}^\dagger$ could be determined for the system. From this the four polarisation elements were determined to be at -7°,37°,85°, and 173°.

Figure 5 shows the polarimetric camera imaging indoors. This consists of a corridor. The doorway at the end is $17.4m$ away from the camera, which is beyond the the parallax limit for the camera. A number of polarising elements have been placed on the door. To test the results, the image was also captured using a more conventional method. A single camera (Pointgrey Flea3) was used to image the scene with a rotated linear polariser placed in front of the lens. The DOLP was calculated at a number of regions and found to differ by only 3% for the well illuminated regions. However, the cameras have a poor response in low light, and thresolding on the s_0 channel maybe required to determine the regions were the DOLP image is inaccurate. The lower part of the door and carpet in the image are at the noise limit of the camera. This results with errors the DOLP image were the values are too high. The carpet should be depolarising and have a very low value.

The angle accuracy measurement was checked in the lab by rotating a polariser. The RMS error in the measurement was 4.1°. The cameras suffer for a higher level of noise than a typical machine vision camera. Measured at a static point in the image the variation of the DOLP had a value of ± 0.03. The angle variation was $\pm 0.7^\circ$ due to noise.

Figure 6 shows another indoors image. This scene consists of metal and plastic piping and power leads. The rear wall is $8.8m$ but the power leads are only $5.5m$ from the camera. The lower right hand region contains a linear polarisation sheet. This is shows as a white square in the DOLP image (Figure 6d). The power lead objects in this image are too near to the camera and suffer from a miss-alignment problem due to parallax. This results in false edges although the rest of the image is correct. Here the images no longer correctly align and incorrect Stokes parameters are measured, although objects slowly varying objects such as the wall will still be correct. This suggests that the more strict requirement of the minimum working distance of $7.5m$ in the Camera Design section was better in the case were the edges of objects present a problem.

CONCLUSIONS

In this paper we develop a low cost polarimetric camera that uses multiple action or dashboard cameras. Each camera is completely independent and can record video and audio onto micro-SD cards or be processed live via a computer. A novel SIFT and cross-correlation method was used to align the cameras. This algorithm doubled the effective field of view to 600×600 pixels compared to conventional SIFT image stitching methods. This still does not utilise the full frame of the camera but due to the short focal length of the lens of the camera there is a significant amount of geometric distortion in the images. Further investigation is needed to solve this problem.

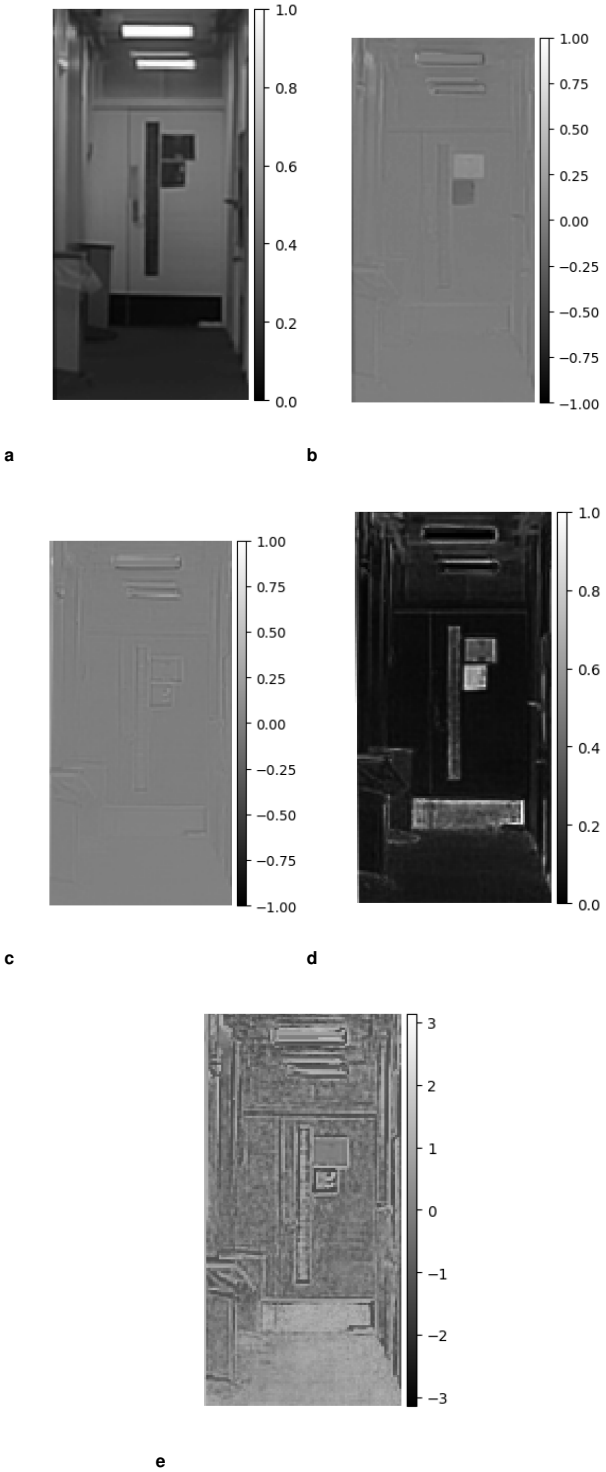


Fig. 5. Indoor result taken with the $N = 4$ camera of a corridor:
(a) s_0; (b) s_1; (c) s_2; (d) DOLP image;(e) ϕ image;

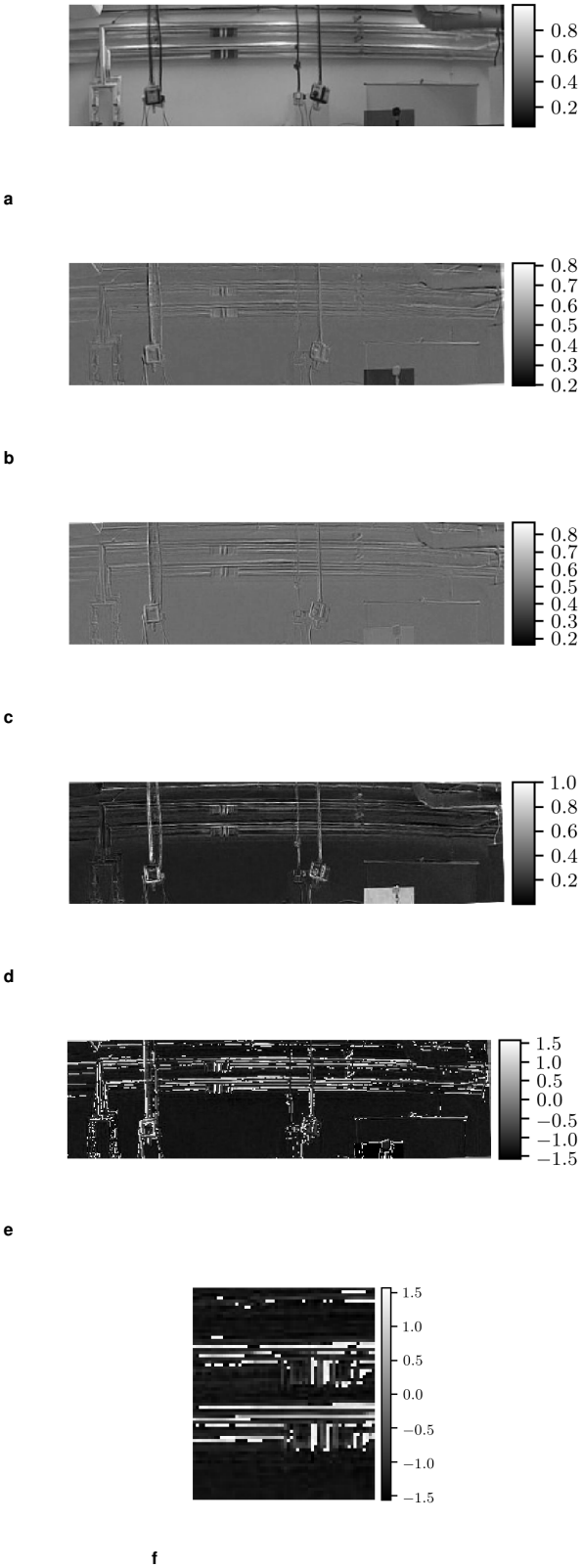


Fig. 6. Indoor result taken with the $N = 4$ camera of a laboratory:
(a) s_0; (b) s_1; (c) s_2; (d) DOLP image;(e) ϕ image;(f) Zoomed in section of the ϕ image showing the foil covered pipes.

Since the cameras are not synchronised, in practice they may be upto a second of time difference between frames. A novel frame syncronisation method using the audio channels was developed to overcome this. Using this, the video could be synchronised to within $\pm 13\mu s$, meaning the process only has to repeated every $70s$.

To model the effect of the cameras' auto-gain, Monte-Carlo simulated images were generated. The error depends on the amount of linearly polarised light within the image. It has a mean value of 2.7% and a maximum value of 7.8% when the $ROLP = 100\%$. This may restrict the application but for a many scenarios the amount of polarisation will not be 100%. The noise effects of the AGC were also modelled. The simulations suggest that the AGC noise does not significantly vary for different amounts of polarisation within the image. Unlike most polarimetric cameras the noise does not increase with the number of measurements due to the light field being divided. This suggests that using more cameras would further improve the image quality provided that a good homography mapping could be produced.

Future work on the cameras is needed to improve the homography mapping between images. This could include more accurate lens models. This would increase the working field of view of the camera, but will make calibration of the device potentially problematic due to the large minimum working distance. This needs further investigation. The noise of the cameras is a limitation as is the poor low light performance. Improving this should improve the errors that appear in the DOLP image when the s_0 values are low and noise dominates.

REFERENCES

1. F. Goudail and P. Réfrégier, "Statistical techniques for target detection in polarization diversity images," Opt. Lett. **26**, 644–646 (2001).
2. L. Meng and J. P. Kerekes, "Adaptive target detection with a polarization-sensitive optical system," Appl. Opt. **50**, 1925–1932 (2011).
3. F. Goudail and M. Boffety, "Fundamental limits of target detection performance in passive polarization imaging," JOSA A **34**, 506–512 (2017).
4. F. Goudail and M. Boffety, "Optimal configuration of static polarization imagers for target detection," JOSA A **33**, 9–16 (2016).
5. W. A. S. Aed El-Saba, Mohammad S. Alam, "Pattern recognition via multispectral, hyperspectral, and polarization-based imaging," (2010).
6. S. Tominaga and A. Kimachi, "Polarization imaging for material classification," Opt. Eng. **47**, 123201 (2008).
7. S. K. Nayar, S. G. Narasimhan, and Y. Y. Schechner, "Polarization-based vision through haze," Appl. Opt. **42**, 511–525 (2003).
8. N. Hautiere, J.-P. Tarel, and D. Aubert, "Towards Fog-Free In-Vehicle Vision Systems through Contrast Restoration," in "Proceedings of the IEEE," (IEEE, Minneapolis, MN, USA, 2007).
9. E. Qu, H. Ju, J. Liang, L. Ren, W. Zhang, Y. Tang, Z. Bai, and Z. Wu, "Real-time image haze removal using an aperture-division polarimetric camera," Appl. Opt. **56**, 942–947 (2017).
10. J. Fade, S. Panigrahi, A. Carré, L. Frein, C. Hamel, F. Bretenaker, H. Ramachandran, and M. Alouini, "Long-range polarimetric imaging through fog." Appl. Opt. **53**, 3854–3865 (2014).
11. F. Liu, J. Xu, P. Han, and X. Shao, "Design of a circular polarization imager for contrast enhancement in rainy conditions," Appl. Opt. **55**, 9242–9249 (2016).
12. J. Tyo, D. Goldstein, D. Chenault, and J. Shaw, "Review of passive imaging polarimetry for remote sensing applications," Appl. Opt. **45**, 5453 (2006).
13. B. Connor, I. Carrie, R. Craig, and J. Parsons, "Discriminative imaging using a LWIR polarimeter," in "SPIE Europe Security and Defence," D. A. Huckridge and R. R. Ebert, eds. (SPIE, 2008), pp. 71130K–11.
14. A.-B. Mahler, D. J. Diner, and R. A. Chipman, "Analysis of static and time-varying polarization errors in the multiangle spectropolarimetric imager," Appl. Opt. **50**, 2080–2087 (2011).
15. N. Gupta, "Development of spectropolarimetric imagers for imaging of desert soils," in "2014 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)," (IEEE, Washington, DC, USA, 2014).
16. W. Sparks, T. A. Germer, J. W. MacKenty, and F. Snik, "Compact and robust method for full Stokes spectropolarimetry," Appl. Opt. **51**, 5495–5511 (2012).
17. A. Jaulin, L. Bigue, and P. Ambs, "High-speed degree-of-polarization imaging with a ferroelectric liquid-crystal modulator," Opt. Eng. **47**, 033201 (2008).
18. N. Lefaudeux, N. Lechocinski, S. Breugnot, and P. Clemenceau, "Compact and robust linear stokes polarization camera," Proc. SPIE **6972**, 69720B–69720B (2008).
19. A. Peinado, A. Lizana, C. Iemmi, and J. CAMPOS, "Polarization imaging with enhanced spatial resolution," OPTICS **338**, 95–100 (2015).
20. E. DeHoog, E. L. Dereniak, H. Luo, J. Schiewgerling, K. Oka, and M. Kudenov, "Compact and miniature snapshot imaging polarimeter," Appl. Opt. **47**, 4413–4417 (2008).
21. R. Perkins, T. York, and V. Gruev, "CCD polarization imaging sensor with aluminum nanowire optical filters," Opt. Express **18**, 19087–19094 (2010).
22. H. Zhao, N. Li, and Y. Zhang, "Polarization calibration with large apertures in full field of view for a full Stokes imaging polarimeter based on liquid-crystal variable retarders," Appl. Opt. **52**, 1284–1292 (2013).
23. F. Goudail and A. Beniere, "Estimation precision of the degree of linear polarization and of the angle of polarization in the presence of different sources of noise," Appl. Opt. **49**, 683–693 (2010).
24. J. S. Tyo, "Optimum linear combination strategy for an N-channel polarization-sensitive imaging or vision system," JOSA A **15**, 359–366 (1998).
25. J. Y. Bouguet, "Camera calibration toolbox for matlab," Tech. rep., Computational Vision at the California Institute of Technology (2015).
26. R. Hartley and A. Zimmerman, *Multiple View Geometry in Computer Vision Second Edition* (Cambridge University Press, Cambridge, 2004), 2nd ed.
27. M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," Int. J. Of Comput. Vis. **74**, 59–73 (2007).
28. M. A. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Commuincations ACM **24**, 381–395 (1981).
29. P. Birch, N. Bangalore, B. Mitra, S. Rehman, R. Young, and C. Chatwin, "Approximate bandpass and frequency response models of the difference of Gaussian filter," OPTICS **283**, 4942–4948 (2010).
30. S. W. Jorge Nocedal, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering (Springer-Verlag New York, 2006).
31. M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," Opt. Lett. **33**, 156–158 (2008).