Processing time-correlated single photon counting data to acquire range images

M. Umasuthan
A.M. Wallace
J.S. Massa
G.S. Buller
A.C. Walker

Abstract: The processing and analysis are described of range data in a time-of-flight imaging system based on time-correlated single photon counting. The system is capable of acquiring range data accurate to 10 μm at a stand-off distance in the order of 1 m, although this can be varied substantially. It is shown how fitting of the pulsed histogram data by a combination of a symmetric key and polynomial functions can improve the accuracy and robustness of the depth data, in comparison with methods based on upsampling and centroid estimation. The imaging capability of the system is also demonstrated.

1 Introduction

Time-correlated single photon counting (TCSPC) is a statistical sampling technique with single photon detection sensitivity, capable of picosecond timing resolution [1]. This technique offers two significant advantages over previous methods for laser ranging based on time-of-flight [2-5]; very accurate time (and hence distance) resolution and very high sensitivity.

Fig. 1 shows a schematic diagram of our sensor, which is described fully elsewhere [6]. A photograph of the optical head is shown in Fig. 2.

The optical source is a passively Q-switched AlGaAs laser diode (developed at the Ioffe Institute, St. Petersburg, Russia), which emits 10-20 ps pulses of energy = 10μJ at 850 nm and a repetition frequency of up to 25 MHz. These pulses are directed towards the target and returned through a series of wave plates, beam splitters and a suitable objective lens. An alternative, known reference channel is provided by a fixed optical fibre. A fraction of each optical pulse is split off and sent to a trigger avalanche photodiode (APD), which acts as a trigger to a single photon counting (SPC) electronics module. A constant fraction discriminator (CFD) is used to determine the temporal location of this reference pulse, which provides the START input for a time-to-amplitude converter (TAC).

Photons scattered from either the target or reference are detected by an actively quenched single photon avalanche diode (SPAD), which provides the STOP input to the TAC. The elapsed time between the START and STOP pulses fixes the output voltage of the TAC, and hence gives the distance to the target. Using a target and reference to define the distance, rather than using the APD signal as a sole reference, enables us to balance any temporal variation in the SPAD signal channel.

The SPAD, developed at the Polytechnico di Milano, is a key element of our system. Besides the general advantages found in using solid-state devices, when compared with photomultipliers these detectors exhibit superior photon detection efficiency, and faster and cleaner time response [7, 8]. The width of the instrumental response (full width at half maximum, FWHM) obtained from the detectors can be less than 50 ps. The small active area (~ 7 μm diameter) presents a considerable advantage in our application, making possible high spatial resolution and low sensitivity to spurious back scatter. An active quenching circuit (AQC) is used to fully exploit the performance of the detectors [9].

Fig. 3 shows an example of a histogram obtained from the photon counting system and represents the number of detected photons (vertical axis) versus time (horizontal axis). The horizontal axis is defined as sampled channels; each channel corresponds to 2.44 ps. The two peaks correspond to the accumulated target (left) and reference (right) single photon returns. The aim of the data analysis is to obtain the time lag (τ) between the reflected target and reference signals in the histogram of counted photons obtained by the SPC module. The distance between the target and reference d is then cτ/2, where c is the speed of the laser light.

In this paper, we consider how data analysis techniques can significantly improve the time resolution of the TCSPC system.

2 Formation of pulse histogram

The accumulated target and reference single photon returns are stored in the SPC board memory. The number of counts in each channel c_i follows a Poisson probability distribution with a mean n_i and a variance n_i [1], where n_i is the number of observed counts in...
In addition to the counting uncertainty, the pulse histogram is contaminated with spurious (dark) counts. To minimise this effect, an estimated dark count is subtracted from the pulse histogram channel by channel.

Another source of distortion in the observed pulse histogram is 'pulse pile up'. Reflected photons arrive at the SPAD randomly and follow a Poisson probability distribution in time. For each laser pulse, there is a finite probability that more than one photon will arrive at the detector. In which case the first received photon is registered, causing a 'pile-up' in the lower channels. Coates [10] derived a correction factor which has been applied to correct the 'pile-up' distortion at high photon collection rates.

Let the probability of an event occurring in channel $c_i$ be $p_i$. An event represents the detection of one or more photons. The number of photon counts in channel $c_i$ is then given by the number of cycles in which the first event occurs in that channel and none in any earlier channel, so that $n_i$ is given by

$$n_i = N p_i \prod_{j=1}^{i-1} (1 - p_j)$$

(1)

where $N$ is the total number of cycles of operation. Eqn. 1 can be solved for $p_i$:

$$p_i = \frac{n_i}{N - \sum_{j=1}^{i-1} n_j}$$

(2)

Let the true probability of detecting a single photon in channel $c_i$ be $s_i$. If $s_i$ is assumed to be constant across channel $c_i$, the number of events in which one or more photons are detected in this channel in one cycle form a Poisson probability distribution, so that

$$p_i = s_i \exp(-s_i) + \frac{s_i^2}{2!} \exp(-s_i) + \cdots$$

$$= 1 - \exp(-s_i)$$

(3)

Rearranging eqn. 3 gives

$$s_i = -\ln(1 - p_i)$$

(4)

The corrected value of $n_i$ is given by $N s_i$.

In general, the shapes of the reference and target single photon returns in the histogram are not identical. For single point measurements, it is usually possible (but tedious) to make these similar by optimising the optical set-up. For multipoint or scene measurement (when one or more objects are present), this is impractical since the target response is dependent on several...
factors, notably the reflectivity and angle of the target surface. Therefore, the total number of counts and the shape of the reference and target single photon returns (after normalisation) are not the same.

The ADC in the SPC board quantises the output of the TAC, so that the probability density function \( f_d(t) \) of detected photons stored as a histogram in the SPC board memory (shown in Fig. 3) is discrete. In effect, the stored histogram is formed by convolving the underlying continuous probability density function \( f_d(t) \) with a uniform probability density function \( f_d(t) \), given by

\[
f_n(t) = \frac{1}{q}, \quad \frac{q}{2} \leq t \leq \frac{q}{2}
\]

where \( q \) is the quantisation width. This is then followed by conventional sampling, i.e. \( f_d(t) = f_d(t) \cdot f_d(t) \cdot \delta_d(t) \), where \( \delta_d(t) \) is the impulse function. The characteristic functions of the continuous and discrete probability density functions, \( \mathcal{F}_d(u) \) and \( \mathcal{F}_d(u) \), are related by

\[
\mathcal{F}_d(u) = \sum_{i=-\infty}^{\infty} \mathcal{F}_c(u + i) \sin \left( \frac{q(u + i)}{2} \right)
\]

The time lag \( t_j \) is the channel width. We used a simple correction term \( \psi \) to correct the errors in the time-of-flight measurement. The correction term \( \psi \) is dependent on the measured time interval and is defined as

\[
\psi = \frac{\text{INL}_{\text{min}} + \text{INL}_{\text{max}}}{2}
\]

and the corrected time measurement \( \hat{t}_j \) is given by

\[
\hat{t}_j = (j - k + 0.5 + \psi) \delta t
\]

### 3 Analysis of histogram data

The time lag \( \tau \) between the reference and target single photon returns can be calculated by finding the separation between the two peaks (Fig. 3) in the histogram. This can be accomplished by either

(i) analysing the raw pulse data in the form of histograms.

(ii) analysing the autocorrelation function formed between the target and reference histograms.

In (i) the pulse separation defines the time lag, and hence the distance to the target. In (ii) the autocorrelation function is symmetric about the central axis, and the distance of the peak of either side lobe with respect to the zeroth time lag defines the time separation. However, in general, each side lobe is not symmetric since this only occurs if the shapes of the target and reference single photon returns are identical.

The common approach to estimate the peak position in either case is centroid estimation. Let \( \hat{c} \) be the location of the \( i \)th channel and \( n_i \) be the number of photon counts in \( c_i \). The centroid is then given by

\[
\hat{\tau} = \frac{\sum_i c_i n_i}{\sum_i c_i}
\]

In most cases, a limited number of channels \( c_i \) s in the neighbourhood of the peak are used in calculating the centroid. If the reference and target single photon returns are well separated in the single photon returns histogram, then all the channels containing reference and target single photon returns can be used in the centroid calculation.

Photon count variations caused by noise introduce deviations in \( \tau \) from the true centroid. As the channels \( c_i \) s in the SPC board memory are random variables which follow a Poisson probability density function with mean \( n_i \) and variance \( n_i \), the centroid is also a random variable. The mean and variance of the centroid can be calculated accordingly. In practice, this produces a large error in the measured peak position.

We concentrate on the analysis of the autocorrelation function, although our approach can be applied (with minor changes) to the raw pulsed data. Analysis in the autocorrelation space has two main advantages: first, the computational time is reduced, since measuring the peak position of one peak of either side lobe gives the time separation of the reference and target photon returns; and second, the dark count effect is minimised.

The autocorrelation function formed from the data shown in Fig. 3 is illustrated in Fig. 4.

![Fig. 4 Autocorrelation of raw histogram](image)

Rather than relying on centroid estimation, we present an alternative approach based on parametric fitting using one of a suitable family of functions to fit...
the autocorrelation data. The analytic form of the function defines the peak, and hence the position on the time axis.

4 Parametric fitting of autocorrelation data

Least squares methods produce parameters with a high probability of being correct if a number of critical assumptions are met.

(a) All of the experimental uncertainty is associated with the dependent variables.

(b) The experimental uncertainties of the data can be described by a Gaussian distribution.

(c) The functional form $f(x, a)$ is correct.

(d) There are enough data points to provide a good sampling of the experimental uncertainties.

(e) The observations are independent of each other.

In the case of the histogram data obtained by single photon returns from the reference and target objects, the assumption that the functional form is known is not valid. Using an inappropriate fitting function which seems to describe the data may produce parameters with no physical meaning. In order to overcome this problem, we use a weighted least squares process combined with a model selection approach to analyse the histograms. To assign the correct weights, we first compute the errors in the discrete autocorrelation data on the basis of the assumed Poisson distributed noise in the channels of the histogram.

4.1 Computing error distribution in autocorrelation data

Let $f_d(t)$ be the single photon returns histogram from the reference and the target. The autocorrelation function is defined by

$$\sum_{t=-\infty}^{+\infty} f_d(t) f_d(t-\tau)$$

where $\tau$ is the time lag.

The number of counts in each channel of the histogram follows a Poisson distribution. For channels containing more than a few counts ($>30$), this can be approximated by a Gaussian distribution with an expectation value equal to the number of counts and a standard deviation equal to the square root of the number of counts in each channel.

For channel $c_k$

$$n_k = \overline{n_k} + \delta n_k$$

where $\delta n_k$ is a random variable with $\overline{\delta n_k} = 0$, and $\sigma_{\delta n_k} = \sqrt{\overline{n_k}}$.

The Fourier coefficient $N_j$ is obtained from

$$N_j = \overline{N_j} + \delta N_j$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} n_k e^{ikj}$$

Separating the real and imaginary parts of eqn. 15 gives the following set of equations [12]:

$$\overline{N_j} = \frac{1}{N} \sum_{k=0}^{N-1} \overline{n_k} \cos \left( \frac{2\pi k j}{N} \right)$$

$$\delta N_j = \frac{1}{N} \sum_{k=0}^{N-1} \delta n_k \cos \left( \frac{2\pi k j}{N} \right)$$

The variances of the Fourier coefficients $\sigma^2_{N_j}$ are given by

$$\sigma^2_{N_j} = \frac{1}{2N} \left( \delta N_j + \delta N_j^* \right)^2$$

The above set of equations can be simplified by using the relation $\delta n_k = \delta n_k$ (Poisson distribution of the counting channels). This gives

$$\sigma^2_{N_j} = \frac{1}{2} \delta N_j^2$$

The autocorrelation function is given by the inverse Fourier transform of $F(u)$, respectively, eqns. 22 and 23 can be written as

$$\sigma^2_{N_j} = \frac{1}{2} \delta N_j^2$$

In terms of the functions $f(t)$, $F(u)$ (the histogram and its Fourier transform, respectively), eqns. 22 and 23 can be written as

$$\sigma^2_{N_j} = \frac{1}{2} \delta N_j^2$$

5 Selection of best fitting model

A conceptual representation of the observed histogram is by a probabilistic density function (PDF). As this underlying PDF is unknown, a nearest representation to the underlying "true PDF" is chosen. This model is referred to as the operating model. Ideally, this operating model could be obtained by modelling the single photon avalanche diode. However, in practice, it is difficult to obtain sufficient information to fully characterise this, and we can only specify the family of models to which the operating model belongs. The size
of the family of models is determined by the number of independent parameters. In principle, the accuracy of the estimation improves when more data are available relative to the parameters.

In order to select appropriate functions, we assume a general family of models which is consistent with the data, given by

\[ g(l, \zeta) = h(l, \theta) \left( 1 + \sum_{j=1}^{m} a_j p_j(l) \right) / \beta \]  

(28)

where \( h(\cdot) \) is the parametric key function; \( p_j(\cdot) \) are polynomials; \( l \) is the time lag of the autocorrelation function; and \( \beta \) is the normalising function of the parameters.

The key function is selected from a family of PDFs. Currently, we are concentrating on the Lorentzian PDF, although we have also used a Gaussian PDF in our experiments.

The Lorentzian PDF is given by

\[ g(l) = \frac{1}{(l - l_0)^2 + \gamma^2} \]  

(29)

where \( l_0 \) is the position of the peak and \( \gamma \) is a measure of the width of the PDF \( (2\gamma = \text{FWHM}) \).

Since the side lobe of the autocorrelation function is not generally symmetric, the parametric key function is multiplied by a polynomial function \( p(l) \). This is selected from the Hermite and Laguerre polynomials [13]. Together with the Lorentzian PDF, this produces a model family that is general enough to fit the empirically observed form of the received data. We have also considered simple polynomials, but these are less effective. The \( r \)th Hermite polynomial is defined by

\[ H_r(l) = (-1)^r e^{\frac{l^2}{2}} D^r e^{-\frac{l^2}{2}} \]  

(30)

where \( D \) is the differentiation operator. The first few terms of the Hermite polynomials are given by

\[ H_0(l) = 1; \quad H_1(l) = l; \quad H_2(l) = l^2 - 1 \]

\[ H_3(l) = l^3 - 3l; \quad H_4(l) = l^4 - 6l + 3 \]  

(31)

The \( r \)th Laguerre polynomial of order \( a \), \( L_a^r(l) \) is given by

\[ \sum_{j=0}^{r} \frac{(-1)^j}{j!} \binom{r+a}{r-j} \frac{l^j}{j!} \]  

(32)

The first few terms of the Laguerre polynomial are given by

\[ L_0^0(l) = 1; \quad L_1^0(l) = a - l \]

\[ L_2^0(l) = \frac{1}{2!} a(a + 1) - (a + 1)l + \frac{1}{2!} l^2 \text{ etc.} \]  

(33)

To select the correct model to fit the data, each combination of parametric key and polynomial functions is tried in turn. The best key and polynomial function parameters are determined by a maximum likelihood estimate (MLE), which uses the error distribution derived in Section 4.1. The iterative fitting algorithm employed is as follows.

Algorithm 1 Operational model fitting

1 repeat

\( \triangleright \) Fit the parametric key function to the side lobe of the autocorrelation function of the target and reference single photon returns histogram.

2 until (convergence)

6 Experimental evaluation of processing methods

For comparison, we have evaluated the performance of the depth measurement system using three different approaches to processing the histogram of reference and target single photon returns.

(i) Computation of the centroid of the side lobe in the original autocorrelated data.

(ii) Computation of the centroid of the side lobe of the reconstructed autocorrelated data following sub-channel reconstruction of the histogram.

(iii) Model fitting using the key and polynomial functions of the reconstructed autocorrelated data following sub-channel reconstruction of the histogram.

We used a 14\( \mu \)m diameter SPAD, with the laser operating at 25MHz and a count rate of about 800kHz. A metal target was mounted on a micropositioner 1.2m away from the sensor. The resolution of the micropositioner was 10\( \mu \)m. A set of 20 measurements was taken, with a collection time of 20s for each measurement. The target was then moved in steps of 20\( \mu \)m towards the sensor. We used a gauge unit (touch probe, Mercer 122) to verify the actual micropositioner movement, and 20 measurements were made at each new target position. This was repeated 10 times in order to get 11 sets of 20 measurements. We then moved the target back to the initial position and repeated the experiment for different step sizes of 30\( \mu \), 40\( \mu \) and 50\( \mu \). The mean of each set was used as the true measurement of the sensor against the ground truth measured by the gauge unit.

![Fig. 5 Comparison of output from three algorithms](image)

Fig. 5 shows a comparison of the results obtained by each algorithm for the same set of data. In this case, a fixed threshold of 100 was used to eliminate the dark counts in the histogram before calculating the autocor-
relation function. In order to calculate the centroid, we used a 50% FWHM threshold of the autocorrelation peak. As stated above, the results obtained by algorithms 1 and 2 depend critically on this threshold, unless the pulse histogram is optimised by fine-tuning the optics and the laser driver. In contrast, the results from algorithm 3 are consistent.

Fig. 6 shows the output of the fitting algorithm; the mean squared error is very low (4.88 x 10^4), but the greatest vertical error in the correlation coefficient occurs on the slopes either side of the peak just above 800 ps. As evident from the experimental data shown in Fig. 7 for example, the horizontal error in peak determination (i.e. time, distance) is very small.

It is possible to obtain better results from algorithms 1 and 2 by varying the threshold (Fig. 7). However, in practice, it is very difficult to set the optimum threshold. In addition, the accuracy obtained by algorithms 1 and 2 is always lower than that obtained by algorithm 3 (Figs. 5 and 7).

In summary, there is good agreement between the processed range and the ground truth determined by the micropositioner. The accuracy is defined as the closeness of the mean measurement between the sensor and the ground truth, and is less than 10μm. The repeatability refers to the closeness of the agreement between successive measurements of the same measurement and can be estimated from the standard deviation, which is 15μm.

Figs. 9 and 10 show range images of a metal object and a toy zebra obtained by the TOF sensor, respectively. Fig. 8 shows an intensity image of the scanned objects. The size of the depth images is 80 x 140 pixels.
The measurements were performed at a photon counting rate of approximately 800kHz and a collection time of 0.5s per point. An XY stage was used to scan the object and, to minimise the possible effects of vibration (caused by the driver circuit of the XY stage), a longer delay was used between each successive measurement. Using one measurement only at each point, we have estimated that the uncertainty associated with each point on the metal object is between $90-150\mu$m, which is higher than the uncertainty associated with repeated single point measurements. This was most probably due to the large variation in the number of collected photons returned from the target with non-uniform reflectivity. However, no exact ground truth was available in this case. We are optimistic that further work on the scanning mechanism and adaptive processing of the received histogram data may substantially improve on this uncertainty.

7 Conclusions

We have demonstrated that time-correlated single photon counting can provide very accurate range data measurement in the order of 10–20μm in reasonable time, in the order of 1s per point. Furthermore, there is a speed/accuracy trade-off; much faster acquisition times are possible with reductions in the accuracy of the range data. We have also demonstrated that the system can be used to acquire dense, occlusion-free range images.

We have also compared different methods to process the reference and target single photon returns histograms. Using an autocorrelation function reconstructed by upsampling improves the basic centroid method, but this is still dependent on heuristic setting of the channel threshold to compute the time separation between reference and target single photon returns. In contrast, fitting of the autocorrelation data by a combination of a symmetric Lorentzian key function adjusted by a Laguerre or Hermite polynomial results in a more accurate estimate of the time separation, which is not sensitive to such thresholds.

8 Acknowledgments

This work forms part of a collaboration between Heriot-Watt University, British Aerospace - Sowerby Research Centre and Edinburgh Instruments Ltd., and has been supported financially by the UK Engineering and Physical Sciences Research Council through their Integrated Machine Vision initiative.

9 References