Using Spread Spectrum Transform for Fast and Robust Simultaneous Measurement in Active Sensors with Multiple Emitters

Anders la Cour-Harbo and Jakob Stoustrup
Department of Control Engineering
Aalborg University
Fredrik Bajers Vej 7C
DK-9220 Aalborg East, Denmark
acle@control.auc.dk, jakob@control.auc.dk

Abstract—We present a signal processing algorithm for making robust and simultaneous measurements in an active sensor, which has one or more emitters and a receiver, and which employs some sort of signal processing hardware. Robustness means low sensitivity to time and frequency localized disturbances, and to white noise. This is achieved partly by using an orthogonal spread spectrum transform for modulating the signals transmitted from the emitters to the receiver, and partly by using a number of transmission channels. The method is fast since the signals are short, and since the method does not rely on previously transmitted signals. This also means that only few calculations are needed. Furthermore, the suggested spread spectrum transform has a low complexity, a high numerical stability, and is easily implemented in simple signal processing hardware. The presented method is therefore suitable for low-cost active sensors.

I. INTRODUCTION

A typical low-cost active sensor transmits a harmonic signal at a particular frequency, and filters the received signal to obtain the energy at the chosen frequency. The energy is the output of the sensing process and is used for various purposes. In many cases a thresholding is applied to this quantity to determine if the transmitted signal is present in the received signal, in which case it signifies some type of occurrence in the proximity of the sensor.

An obvious drawback of using a harmonic signal is that such signals are common in most environments. They appear as electrical, optical, acoustic, mechanical signals, and it is futile to isolate the sensor from all harmonic signal sources. Another less obvious problem with the above described sensor is that it does not respond well to short and powerful burst of noise. If the sensor is damaged, short-circuited, or subjected to extreme signal conditions (like a flash close to an photosensitive diode) a typical low-cost sensor might easily interpret the situation wrongly, as it only cares about the energy at a certain frequency, and not the circumstances that caused the presence of energy. For instance, heavy white noise contains a lot of energy at all frequencies, but should not be interpreted as a successful transmission of a harmonic signal.

A. Background on the Presented Solution

This paper presents a method for increasing the robustness of a sensor by purely algorithmic means. Such an approach assumes the presence of digital signal processing hardware in the sensor. An algorithm implemented in the hardware handles the generation of the signals to be transmitted as well as the post-processing of the signals after transmission. This paper presents an algorithm which can handle these tasks in a way which makes the sensing process fast and insensitive to common types of noise.

The aim is to have a method, which is fast and robust while still being suitable for low-cost signal processing hardware. In this context fast means short time from occurrence to detection. Optimally, this delay is (almost) equal to the sampling frequency, which means that the sensor must respond to an occurrence based on just a single sample. This is in contrast to the desire for robustness, which calls for several samples to verify that the believed occurrence is real and not just a burst of noise.

On top of this it is also desirable to have a fairly simple and straightforward method as this reduces the requirements to the signal processor. Although the ratio between computing power and cost has increased rapidly in the past years, DSPs and the like are still a comparatively expensive component in low-cost active sensors. Consequently, it is rather important that an active sensor algorithm remains simple in all aspects of implementation. Finally, the spread spectrum modulation should produce binary sequences to avoid the need for a digital-to-analog converter.

B. The Presented Solution

In the present effort we construct a simple and generic method that can be applied in many types of active sensors with one or more emitters, and which complies with the requirements presented in the previous section. We suggest a method where it is easy to adjust the trade-off between response time and robustness, and which has a low program complexity, high numerical stability, and is easy to implement. We also want the method to be applicable to systems with multiple emitters and receivers.

The basic idea of the suggested algorithm is to combine spread spectrum modulation with a multiplicity of transmission channels, i.e. the channels are separated in the 'spread spectrum domain' or code domain (just like radio broadcasting is a separation in the frequency domain). To perform the sensing a number of emitters simultaneously emits a single ping, like a sonar. This ping is actually short, typically 16 to 64 samples long, spread spectrum sequences. Each emitter has its own sequence, and all the sequences are different from each other and each represents a channel. Each receiver then receives a mix of all the signals, which are processed to determine if there has been an occur-
rences (based on the intensity of the ping in each channel) to which the sensor should respond. The method reported in this paper applies to the receivers individually, so in the following only one receiver is considered.

Since the channels are not separated in neither the time nor the frequency domain any noise occurrences localized in time or frequency will affect all channels approximately evenly. The idea is then to use only a few channels for actually transmission while the remaining channels are used for detecting noise. That is, no ping is emitted into these channels. The algorithm is thereby capable of detecting when the noise is such that the transmission is corrupted beyond recognition. At the same time the modulation also increases the immunity of the short signals transmitted in the individual channels to time and frequency localized noise. By changing the length of the short signals it is possible to easily change the trade-off between response time and robustness.

II. Method for Spread Spectrum Modulation in Multiple Channels

One of the key elements in the algorithm is the modulation method. We propose to use the Rudin-Shapiro transform (RST) for this purpose, since it has a series of useful properties (see subsections II-A and II-B).

The sensing starts with a set of very simple digital signals to indicate which channels are used for pinging. The signals are zero sequences with a 1 at the location of the chosen channel. In Fig. 1 three such signals are shown. An RST of such an almost vanishing signal produces a binary spread spectrum signal with a number of samples equal to the original sequence. These signals are transmitted by a number of emitters to the receiver through the sensor environment, and the received signal is now demodulated with the RST. If the receiver behaves in a linear fashion and because of the orthogonality properties of the RST the entries in the demodulated signal corresponding to the chosen transmission channels now holds the energy of the received pings (yi holds the energy from the first emitter, y2 holds the energy from the second emitter, and so on) while the remaining entries up to yN are zero.

In any real life application there is obviously noise present, and thus the remaining entries are not zero. However, they always remain unaffected by the transmission from the emitters, and can therefore be used to determine properties of the current noise. This information can in turn be used to validate the quantities obtained from the transmission from the emitters. Two such methods are reported in subsection III-A and III-B, respectively.

A. Rudin-Shapiro Transform

The RST is defined through the remarkable Rudin-Shapiro polynomials, introduced in 1951 by H. S. Shapiro in his master's thesis [1], and published in 1959 by Rudin [2]. The RST is a symmetric, orthogonal Hadamard $2^J \times 2^J$ matrix, for $J \in \mathbb{N}$, where each row consists of ±1's and is the coefficients of a flat polynomial. Hence these binary sequences constitutes an orthonormal basis for $\mathbb{R}^{2^J}$ with basis vectors consisting of spread spectrum sequences of ±1's. In figure 2 is an example of such a sequence for $J=2$.

Since the Rudin-Shapiro transform decomposes a signal into a linear combination of spread spectrum sequences the transform coefficients of any time or frequency localized signal will always have a slow decay. This means that if a signal is subjected to a time or frequency localized burst of noise and subsequently Rudin-Shapiro transformed the energy of the burst is distributed fairly evenly among all the transform coefficients, while the total energy of the burst is unchanged because the transform is orthogonal. Effectively, the impact of the burst of noise is changed from being local and big to global and small.

The unitarity of the transform also allows easy separation of several signals. If, say, three signals of different and unknown amplitude are mixed (as a linear combination) into x, and the three signals are scaled versions of the first three rows in the RS matrix, an RST will immediately reveal this; the first three coefficients of the transformed signal $y = \text{RST}(x)$ are the amplitudes, while the remaining coefficients are zero. Consequently,
an orthogonal transform provides a number of channels equal to its matrix dimension through which information can be transmitted independently. Also, note that a normally distributed stochastic process is still normally distributed after an orthogonal transform.

B. Implementation of the Transform

The properties of the RST presented above makes the transform interesting in a number of applications. But for the transform to be really useful a fast version is needed. Multiplication by a matrix is an $O(N^2)$ operation, which for longer signal can become a problem in real-time applications. Fortunately, there exists a very simple and fast RST implementation.

The fast RST of size $2^J \times 2^J$ consists of $J$ intermediate steps which are applied sequentially. Each intermediate step is a filtering with a low pass filter $[1 \ 1]$ and a high pass filters $[1 \ -1]$. However, the filters are not stationary. The minus 'moves' between the four filter taps as the filtering occurs, effectively making the low and high pass filters swap places and do time reversals. This continues interchange of the low and high pass filters can also be regarded as a simple explanation for the spread spectrum property of the transform. To get rid of the square root one can also be regarded as a simple explanation for the spread spectrum property of the transform. To get rid of the square root $1/\sqrt{2}$ can be applied as $1/2$ after every other intermediate step (since the steps are linear). Consequently, a fast RST requires exactly $J \cdot 2^J$ additions and subtractions, and $J/2$ multiplications by $1/2$, thereby reducing the complexity to $O(N \log N)$.

Finally, note that a symmetric, orthogonal transform is its own inverse; once the RST is implemented, so is its inverse.

III. METHODS FOR VALIDATION OF MEASUREMENTS

We report two different methods for using the extra channels to increase the robustness of the sensor. The first method is a statistical model for balancing the probability of making a false positive (FP) decision, i.e. responding to a non-existing occurrence, and a false negative (FN) decision, i.e. not responding to an occurrence. The second method is an ad hoc approach, which attempts to address some of the problems in the first method. However, while the first method is theoretically founded, the second method is to some extend empirical founded. It is important to note that both methods are designed such that, in order to lower the response time, a single ping is enough for validation.

Both methods are based on the SNR, that is the relations between the intensity in the pinged channels and the intensity in remaining channels. To have the simplest method the pinged channels are treated individually. Therefore, in this section we assume $y_0$ to be the pinged channel, and $y_1$ through $y_N$ to be all the non-pinged channels. The basic observation for both methods is that a good estimate of whether the emitted ping is present in the received signal is the ratio between signal and noise, i.e. a test on the form

$$T(\alpha) : y_0^2 > \alpha \frac{1}{N} \sum_{k=1}^{N} y_k^2$$

(1)

where $N$ is the number of test channels. Basically, if $T$ is fulfilled, it is assumed that a ping was received, and if $T$ is not fulfilled, it is conversely concluded that no ping (of sufficient magnitude) was received.

Obviously, the risk of detecting a signal when none was received (an FP decision) decreases with large values of $\alpha$. Vice versa, when $\alpha$ is small, the risk of ignoring a signal that was actually received (an FN decision) is also small.

A. First Validation Method

The purpose of the first method is to determine the optimal threshold $\alpha$. We define an optimal threshold as that value of $\alpha$ for which the probability for a worst-case FN decision based on $T(\alpha)$ equals the probability for an FP decision based on $T(\alpha)$. A worst-case FN decision is understood as an FN decision in the presence of the faintest received signal which should be detected. It is straightforward to modify the approach below in order to meet this compromise with a preference to either a low FP or a low FN probability.

The reader is reminded that if $N$ stochastic variables $\{y_k\}_{k=1}^N$ are normally distributed, $Y_k \in \mathcal{N}(0, 1)$, then

$$\sum_{k=1}^{J} Y_k^2 \in \chi^2(N)$$

is the worst-case SNR, i.e. $R_{max}$ is the lowest detectable signal level of $y_0$. Note that this makes $R_{max}$ the 'real' SNR since $y_0$ is (for the time being) assumed to be deterministic. $R_{max}$ would typically be a design parameter or adjustable by the user of the sensor. Now, the probability distribution function $P_{FP}$ can be computed by

$$P_{FP}(\alpha) = \int_{A} f_{Z_0,Z_2} dz_0 d\bar{z} = \int_{A} \frac{1}{\Gamma(\frac{N}{2})\sqrt{2}} e^{-\frac{z_0^2}{2}} \frac{1}{\Gamma(\frac{N-2}{2})} e^{-\frac{\bar{z}^2}{2}} \frac{1}{2} d\bar{z}$$

(2)

Introducing polar coordinates, the integral in (2) becomes

$$\int_{\theta=0}^{\pi/2} \int_{r=0}^{\infty} \frac{2^{-\frac{N+1}{2}}}{\Gamma(\frac{N}{2})\Gamma(\frac{N-2}{2})} r^{N-1} e^{-r} (\cos \theta + \sin \theta)^{N-2} \theta^{\frac{1}{2}} d\theta d\theta$$

(3)

The double integral can be separated into two single integrals by the substitution $r = \bar{r} (\cos \theta + \sin \theta)$, that is,

$$P_{FP}(\alpha) = K_1 \int_{0}^{\arctan \frac{1}{\alpha}} \left( \frac{\cos \theta + \sin \theta}{\cos \theta \sin \theta} \right)^{N-2} d\theta$$

(4)

$$K_1 = \frac{2^{-\frac{N+1}{2}}}{\Gamma(\frac{N}{2})\Gamma(\frac{N-2}{2})} \int_{0}^{\infty} r^{N-1} e^{-\frac{r}{\alpha}} d\bar{r}$$

2671
Finally, it can be shown that the standard substitution \( t = \tan \frac{\alpha}{2} \) leads to the following algebraic integrand

\[
P_{FP}(\alpha) = 2 \pi K_1 \int_0^{\sqrt{1+\alpha^2}} \frac{t^{N/2}}{(2 - (t - 1)^2)^{N/2}} \sqrt{1 - t^2} dt.
\]

Even though the integral in (5) is algebraic, it can not be resolved analytically. However, numerical experiments show that e.g. an adaptive recursive Newton Cotes 8 panel rule performs better on (5) than on (4). The resulting probability function for \( N = 13 \) is shown in Fig. 3 (\( N = 13 \) is chosen to match the experimental signals in Section IV).

Exploiting the probability function \( P_{FP}(\alpha) \) which can be evaluated numerically by (5) it is straightforward to get a calibrating curve for \( \alpha \) under the constraint that \( P_{FP} = P_{FN} \). To that end, we start with a value of \( \alpha \), and numerically determine \( P_{FP}(\alpha) \). Then the inverse of the \( \chi^2(N) \) distribution function applied to \( P_{FP}(\alpha) \) yields the ratio between \( R_{max} \) and \( \alpha \). This relationship is shown in Fig. 4. While this curve is valid only for deterministic \( \gamma_0 \), the same type of curve can be generated under the assumption that \( \gamma_0 \) is under influence of a stochastic process. In that case the curve is asymptotic to a horizontal line for small values of \( R_{max} \) (see [3]).

For this test to be useful we must have \( \beta \ll 1 \), and then \( \beta \) is important only when \( \gamma_0 \) is of a magnitude at least equal to the magnitude of the noise. A big \( \beta \) makes the inequality more sensitive to heavy noise burst (where \( \gamma_0 \) as well as \( \gamma_1 \) through \( \gamma_N \) are large). But the limit of \( \gamma_0 \) for which the inequality always evaluates to false is lowered with a big \( \beta \). In a low noise scenario the sum is negligible, and the inequality is then false when \( \beta > \alpha/\gamma_0 \), approximately. Thus, a good choice is \( \beta = \alpha/\gamma_{max} \), where \( \gamma_{max} \) is the dynamic range of \( \gamma_0 \). The \( \alpha \) factor still adjusts how much we want to trust the measurements classified as useful, and is thus very application specific.

One can also determine the corresponding \( P_{FP} = P_{FN} \) curve, which is seen in Fig. 5. Note, that this calibration curve has a left, horizontal asymptote corresponding to the limiting case \( P_{FP} = P_{FN} = \frac{1}{2} \) as the SNR goes to zero. In a doubly logarithmic plot, it also has a right asymptote, which in fact is the curve \( \alpha = (\beta \sigma \cdot \text{SNR})^{-1} \).
IV. Results

Some result of applying the spread spectrum modulation and validation methods on real data is presented in this section. We have tested the methods in a physical setup, which is described in the next subsection. The outcome of this is presented in the following subsections.

A. Test Setup and Test Signals

The test data is obtained with three infra red diodes and one infra red photodiode. The are all located in the same plane and facing the same way, such that transmission from emitters to receiver occurs by reflection. The diode driver electronics are connected to a PC via 12 bit sampling boards. The algorithms are implemented in ANSI C for real time execution. The RST modulation produces 16 samples per ping (allowing for 16 channels), and these samples are continuously transmitted at 2.6 kHz in 4 μs wide pulses. The ping rate is thus 163 Hz. The test data intentionally has a low SNR, since we want to investigate the behavior in noise conditions.

Two experiments have been conducted, generating two sets of test signals. Each signal consists of 1060 pings (a good 6 seconds) recorded with an objected moving in front of the diodes to reflect the light. Three emitters are used, so three channels are occupied by transmitted signals. The remaining 13 channels are noise channels. Since the methods applies to the channels individually, only one channel is examined. The other channels can be treated equally. Once the first channel is evaluated the remaining pinged channels can be evaluated at almost no extra cost, since the sum takes the major part of the computation.

In the first test signal set only random noise is present. Any transient-like structure is thus a result of ‘valid’ occurrences, and not noise. In the second test signal set a noise occurrence have been generated by touching the driver electronics for the receiver with a screw driver at its most sensitive point. The screw driver essentially works as an antenna and the result is noise occurrences as seen in the second test signal.

The two validation methods use a threshold on the ratio between signal and noise to classify the GM. When adapted to the test signals at hand the first validation method is

\[ T(\alpha) : \Theta(y) = \frac{y_0^2}{\sum_{k=1}^{15} y_k^2} > \alpha \]

while the second method is

\[ \tilde{T}(\alpha) : \tilde{\Theta}(y) = \frac{y_0^2}{\sum_{i=1}^{15} y_i^2 + \beta |y_0|^2} > \alpha . \]

B. Applying the First Validation Method

The first thing to do is determine worst-case SNR, i.e. the weakest detectable signal compared to the expected random noise level. The weakest signal is chosen to be \( y_{\text{low}} = 15 \) (this depends on the maximum distance at which a given object should be detectable). Since the variance \( \sigma^2 \) is approximately 9.1, \( R_{\text{max}} = y_{\text{low}}/\sigma = 4.8 \). Using the curve for stochastic \( y_0 \) (this differs a little from the curve in Fig. 4) this yields approximately \( \alpha = 0.5 \) for which \( P_{FP} = P_{FN} \approx 0.02 \).

The first test signal subjected to the first validation methods is shown in Fig. 6.

The validation behaves as expected. For the ‘no signal’ part in the beginning \( T(\alpha) \) is false for almost all samples (approximately 1 out of 50 is expected to be accepted as a useful GM). When the signal level approaches \( y_{\text{low}} \), which equals 15 and is shown with a dashed line, the validation shifts in favor increasingly more useful measurements. Again the fraction of measurements validated incorrectly is 1/50 for signal level close to \( y_{\text{low}} \). Note that the point in time (here measured in samples) at which more measurements are considered useful than not useful is the same as where the average level of the signal is \( y_{\text{low}}/2 \). This happens around sample number 250. This corresponds with the notion that if the signal itself was used for validation and \( P_{FP} \) should equal \( P_{FN} \) for the weakest detectable signal the threshold should be half the weakest detectable signal level.

Now, applying the validation to the other test signal and using
the same parameters the result is a less gratifying. A large portion of the samples are evidently classified as useful although they are not. It is worth noting that the validation does not fail because there is no difference significant difference between $\Theta(y)$ for the useful and useless parts of the third test signal. It fails because the threshold is wrong. However, the threshold has been determined such that it complies with our notion of proper behavior in a random noise scenario, and accordingly it works fine for the signal without transients. This problem can be solved easily by increasing the threshold to around 20, which seems to separate nicely the useful and useless parts in the third test signal. However, the result of simply increasing the threshold is that virtually all the samples in the first test signal are classified as useless. Evidently, another validation method is needed to handle this problem.

C. Applying the Second Validation Method

This is why the second validation method is relevant. The introduction of this second method is solely an attempt to handle the transient in a proper manner while at the same time responding to random noise exactly as the previous method.

The first step is to determine $\alpha$. Using the $P_{FP} = P_{FN}$ curve for the second method, see Fig. 5, yields a value a little smaller than for the first validation methods, namely $\alpha = 0.6$. The $\beta$ is determined according to the guidelines given in Section III-B with a maximum value of $\gamma$ set to approximately 100. This gives $\beta = 0.017$. To ensure a not too large value $\beta = 0.012$ is used. The result is shown in Fig. 7. It is clear that the results in the first test signal is approximately the same as in the first validation method, but the second test signal is now in general classified correctly. More of the transients can be 'caught' by increasing $\beta$, but as described previously this lowers the maximal acceptable value of $\gamma_0$. In this case experiments have shown that increasing $\beta$ to 0.016 will produce a very good result. However, if this value is chosen, new $\gamma_0$ signals must not exceed 100, or they will be classified as useless.

A third option has been introduced in the second validation method. If the signal $\gamma_0$ is too small (here that is below $\gamma_{low}/2$) it is not attempted classified, because the second validation method can become unstable for small $\gamma_0$.

V. DISCUSSIONS

We have presented a method for low-cost active sensor which provides increased robustness in comparison to the traditional method of using harmonic signals and followed by thresholding for validation. The use of spread spectrum signals provides some of the robustness, and the estimation of measurement validity by means of channels without transmission provides some of the robustness. Moreover, the method is easy to implement and is well-suited for low-cost signal processing hardware, even fixed point microprocessor.

These nice properties does not come without a cost, however. Using only very little information to determine the validity the method is bound to fail occasionally, and since the adjustment of parameters have not been automated it requires manual adjustment in each applications, and possibly also on a regular basis.

VI. ACKNOWLEDGMENTS

This work is in part supported by the Danish Technical Science Foundation Grant no. 9701481. Thanks are due to LEGO Engineering, Denmark, for assistance with construction of the test setup.

REFERENCES