Cross-Modal Localization Through Mutual Information

Alen Alempijevic, Sarath Kodagoda and Gamini Dissanayake

***Abstract*— Relating information originating from disparate sensors observing a given scene is a challenging task, particularly when an appropriate model of the environment or the behaviour of any particular object within it is not available. One possible strategy to address this task is to examine whether the sensor outputs contain information which can be attributed to a *common cause*. In this paper, we present an approach to localise this embedded common information through an indirect method of estimating mutual information between all signal sources. Ability of L_1 regularization to enforce sparseness of the solution is exploited to identify a subset of signals that are related to each other, from among a large number of sensor outputs. As opposed to the conventional L_2 regularization, the proposed method leads to faster convergence with much reduced spurious associations. Simulation and experimental results are presented to validate the findings.**

I. INTRODUCTION

The world market for sensors and wireless communication technologies is ever growing, prompting the rapid deployment of wireless sensor networks [1]. Therefore, it is not unreasonable to assume that sensors will be omnipresent in the near future. With the presence of large number of sensors and signals, there is a growing interest in cross-modal signal analysis. The objective is not necessarily to geometrically relate the sensors, the emphasis is rather placed on relating parts of the sensor signals. The following fundamental concept in perception is exploited extensively in this paper: motion has in principle, greater power to specify properties of an object than purely spatial information. Thus, relating signals could generally be carried out through comparison of vectors of signals, which have been monitored over time. One important aspect of such signal processing is to *localize* some components of a particular signal to that best correlate with the other signal, which also originated from the same source.

This type of analysis is reported in various fields including, biomedical engineering, climatology, network analysis and economy. In biomedical research, heart rate fluctuations are examined against several interacting physiological mechanisms including visual cortex activity, respiratory rate etc [10] in order to determine the neurological status of infants. In climatology, dynamic weather patterns in a particular location are correlated to synoptic meteorological data gathered over time [13]. In economy, revenue performance of a market is correlated with a large set of economic and social criteria [15].

A. Alempijevic, S.Kodagoda and G. Dissanayake are with ARC Centre of Excellence for Autonomous Systems (CAS), University of Technology, Sydney, Australia a.alempijevic, s.kodagoda, g.dissanayake @cas.edu.au

There a number of techniques that are suitable for detecting the statistical dependence of signals. Techniques such as Canonical Correlation Analysis and Principle Components Analysis rely on correlation, a second order statistic. Alternative non parametric techniques are Kendall's tau, Cross Correlograms, Mutual Information (MI) and Independent Component Analysis. The selected metric is required to identify a non-linear higher (than second) order of statistical dependence between signals. The measure of statistical dependence should be valid without any assumptions of an underlying probability density function and should be extendible to high dimensionality of input signals. Mutual information is identified as the most promising metric, fulfilling all requirements.

The methods for mutual information (MI) estimation can be classified into two broad categories, based on whether mutual information is computed directly or the condition for maximum MI is obtained indirectly through an optimization process that does not involve computing MI [2], [7]. The most natural way of estimating MI via the direct method is to use a nonparametric density estimator together with the theoretical expression for entropy. However, the definition of entropy requires an integration of the underlying PDF over the set of all possible outcomes. In practice, there is no closed form solution for this integral. Combining the nonparametric density estimator with an approximation of theoretical entropy has been widely described in the literature to overcome this problem [16]. However, this requires pair wise comparisons of all permutations of input signals to find the most informative statistically dependent pairings, which is not feasible for large number of signals, such as images.

The indirect MI estimation method determines the most mutually informative signal pairings through mapping of the signals into a two dimensional space. The key to obtaining the most informative mapping is in a technique that computes the effect of the mapping parameters on the information content in the lower dimensional space. Fisher et. al. [8] demonstrate a linear mapping of the signals that maximise MI by defining an objective function that operates on the resulting two dimensional space.

This paper builds upon Fisher's work [8] and our previous research on indirect MI estimation [2] by introducing the L_1 norm to obtain a sparse linear mapping. L_1 norm has found extensive use recently in solving convex optimisation problems from arbitrary signals estimated from incomplete set of measurements corrupted by noise [5] and also exhibits a very useful property, which is the preservation of the sparsity of the relationship between the multidimensional random variables. The L_1 norm as a penalty function on

the magnitudes of the mapping coefficients is shown to be suited to the applications examined in this paper where the mutually informative signals are usually embedded in a large number of non informative signals.

The remainder of this document is organised as follows, Section II outlines an indirect estimation algorithm for MI. Section III describes the process of finding the maximum MI with L_1 penalty norm and optimization parameters. Experimental results are presented in Section IV. Section V concludes the paper providing future research directions.

II. Indirect Estimation of Mutual Information Through Non-Linear Mappings

Mutual information between two random vectors X_1, X_2 can be defined as follows.

$$I(X_1; X_2) = H(X_1) + H(X_2) - H(X_1, X_2) \quad (1)$$

Where, $H(X_1)$ and $H(X_2)$ are the entropies of X_1 and X_2 respectively, $H(X_1, X_2)$ is the joint entropy term. Direct estimation of MI requires calculation of entropy terms in (1). Entropy $H(X_1)$, also referred to as Shannon's entropy of random variable X_1 with density $p(x_1)$ is given by,

$$H(X_1) = -\int_{\Omega} p(x_1)\log(p(x_1))dx_1 \quad (2)$$

where Ω is the set of possible outcomes.

There are two distinctive problems that need addressing when calculating entropy in this form, firstly calculating the underlying unknown PDF of the random variable to obtain $p(x_1)$ over the entire space Ω, and second, the integration over the set of all possible outcomes. Both are addressed through indirect estimation.

Mutual information between two high dimensional signals X_1 and X_2 can be indirectly estimated by mapping the signals into a lower dimensional space, by exploiting the data processing inequality [6] that defines lower bounds on mutual information. The inequality states

$$I(g(\alpha_1, X_1); g(\alpha_2, X_2)) \leq I(X_1; X_2) \quad (3)$$

for any random vectors X_1 and X_2 and any function $g(\alpha, \cdot)$ defined on the range of X_1 and X_2 respectively. The generality of the data processing inequality implies that there are no constraints on the choice of transformations $g(\cdot)$. Furthermore, as the functions $g(\alpha, \cdot)$ map the input data into a lower dimensional space, computing the information content $I(g(\alpha_1, X_1); g(\alpha_2, X_2))$ is significantly easier.

The mappings $Y_1 = g(\alpha_1, X_1)$ and $Y_2 = g(\alpha_2, X_2)$ can be achieved through any differentiable function, such as hyperbolic tangent [11] or multiple layer perceptrons [8]. However, linear projections are preferred due to the fact that the linear projection coefficients themselves can be used as a measure of MI of each individual signal in random vectors X_1, X_2 to the resulting lower dimensional Y_1, Y_2 mutual information. We now present how to select the parameters of linear mappings $Y_1 = \alpha_1 X_1$ and $Y_2 = \alpha_2 X_2$, thus, selecting subset of the most mutual informative signals from sets of signals X_1 and X_2 without the need to estimate MI on all permutations of signal sets.

III. Optimization of Mappings via Information Maximisation Principle

Finding the optimal projections α_1 and α_2 would require solving a complex non-linear optimization problem. It is generally not feasible to obtain a closed form solution to this problem without numerical methods such as Powell's direction set method [3]. However, the high cost of computing MI, together with the fact that the parameter vector α is in the dimension of the input signals in the case of a linear map makes direct optimization intractable.

An entropy estimation measure proposed by Fisher et. al. [8] allows for obtaining the gradient of the measure with respect to the mappings parameters. They proposed an unsupervised learning method by which the mappings $g_1(\cdot)$ and $g_2(\cdot)$ can be estimated indirectly, without computing mutual information. The maximisation of MI is achieved by maximising the entropies $H(Y_1)$ and $H(Y_2)$ and minimising the joint entropy, $H(Y_1, Y_2)$ in (1). The entropies $H(Y_1)$ and $H(Y_2)$ can be maximised by selecting the mapping parameters to make the data on the lower dimensional space resemble a uniform distribution. Likewise, joint entropy $H(Y_1, Y_2)$ can be minimised by selecting the mapping parameters to reflect the joint distribution, (Y_1, Y_2) is furthest away from a uniform distribution.

Thus, maximisation of MI can be achieved by maximising the objective function J,

$$J = J_{Y_1} + J_{Y_2} - J_{Y_{1,2}} \quad (4)$$

where each element of $J_{Y_1}, J_{Y_2}, J_{Y_{1,2}}$ are of the form,

$$\frac{1}{2}\int_{\Omega}\left(f(u) - \hat{f}(y_u)\right)^2 du \quad (5)$$

Where Ω indicates the nonzero region over which the integration is evaluated. Therefore (5) is the integrated square distance between the output distribution (evaluated by a parzen density estimator, $\hat{f}(y_u)$ at a point u over a set of observations y) and the desired output distribution $f(u)$.

It can be shown that the gradient of each element of J with respect to the mappings parameters α can be computed as follows [8].

$$\begin{aligned}\frac{\partial J}{\partial \alpha} &= \frac{\partial J}{\partial \hat{f}}\frac{\partial \hat{f}}{\partial g(\alpha,x)}\frac{\partial g(\alpha,x)}{\partial \alpha} \\ &= -\frac{1}{N}\sum_i \epsilon_i \frac{\partial}{\partial \alpha} g(\alpha, x)\end{aligned}$$

Note that $\frac{\partial g(\alpha,x)}{\partial \alpha}$ is a constant as we have assumed $g(\cdot)$ is a linear projection. The term ϵ_i is [8],

$$\epsilon_i^{(k)} = b_r(y_i^{(k-1)}) - \frac{1}{N}\sum_{j\neq i}\kappa_a(y_i^{(k-1)} - y_j^{(k-1)}, \Sigma) \quad (6)$$

$$b_r(y_i)_j \approx \frac{1}{d}\left(\kappa_a(y_i + \frac{d}{2}, \Sigma)_j - \kappa_a(y_i - \frac{d}{2}, \Sigma)_j\right) \quad (7)$$

$$\kappa_a(y,\Sigma) = G(y,\Sigma) * G^{'}(y,\Sigma) \tag{8}$$

expanding G and $G^{'}$

$$\kappa_a(y,\Sigma) = -\frac{1}{2^{M+1}\pi^{M/2}\sum^{M+2}} exp(-\frac{y^T\Sigma^2 y}{4})y \tag{9}$$

where, $\kappa_a(.)$ is a kernel: a Gaussian PDF with standard deviation of $\Sigma = \sigma^2 I$ is assumed here. y_i symbolises a sample of either Y_1 or Y_2 or the concatenation, $Y_{1,2} = [Y_1; Y_2]$ for $J_{Y_{1,2}}$, M is the dimensionality of the output space and is $M = M_1, M_2$ or $M_1 + M_2$ based on the term of (4) that is considered. The j^{th} element of $b_r(y_i)$ in (7) is defined as $b_r(y_i)_j$, d is the support of the output space and N is the number of samples.

For systems where the dimensionality of the input space N is more than the number of samples n, the mapping can be arbitrary. To obtain a single solution a penalty on the projection co-efficients α_1 and α_2 can be imposed. The minimal energy solution can be obtained by imposing the L_2 penalty while the L_1 norm is shown to lead to the sparsest solution. The fact that the L_1 penalty leads to a vector with fewest nonzero elements for both overdetermined and underdetermined systems has been demonstrated [14].

A. Optimizing Linear Mappings via the L_2 Regularisation

Projection coefficients that maximise the objective function can now be found using the algorithm given in Fig. 1 which includes the update rule (6) for each entropy term (1) and imposition of a L_2 penalty ($L_{2(\alpha_1)}, L_{2(\alpha_2)}$) on the projection coefficients α_1 and α_2

$$J = J_{Y_1} + J_{Y_2} - J_{Y_{1,2}} - \beta\left(L_{2(\alpha_1)} + L_{2(\alpha_2)}\right) \tag{10}$$

where the L_2 penalty is derived from

$$L_{2(\alpha_1)} = \frac{\partial \alpha_1 \alpha_1^T}{\partial \alpha_1} \tag{11}$$

therefore

$$L_{2(\alpha_1)} = 2Y_1 X_1^{-1}\left({X_1}^{-1}\right)^T \tag{12}$$

$$L_{2(\alpha_2)} = 2Y_2 X_2^{-1}\left({X_2}^{-1}\right)^T \tag{13}$$

where X^{-1} is the pseudo inverse of matrix X.

B. Optimizing Linear Mappings via the L_1 Regularisation

The L_2 criterion seeks to spread the energy of α_1 and α_2 over many small valued components, rather than concentrating the energy on a few dominant ones. The applications examined in this paper, requires identifying a few dominant components in the input signal space that are related to each other. Hence, the solution of the parameter vectors α_1 and α_2 should be sparse identifying the minimum number of nonzero elements naturally suggesting the use of the L_1 norm as an appropriate penalty function. In addition, the number of samples and dimensionality of the signals can vary between applications producing an either underdetermined or overdetermined system of equations $Y_1 = \alpha_1 X_1$ and $Y_2 = \alpha_2 X_2$. The L_1 norm performs equally well as the L_2 norm on overdetermined system of equations while outperforming L_2 norm for underdetermined problems [9] especially where the solution is expected to have fewer non zeros than $1/8$ of the number of equations.

The update equation for the gradient descent method when using the L_1 penalty is

$$J = J_{Y_1} + J_{Y_2} - J_{Y_{1,2}} - \beta\left(L_{1(\alpha_1)} + L_{1(\alpha_2)}\right) \tag{14}$$

The equations for the L_1 norm penalty are derived

$$\begin{aligned} &\min \|\alpha_1\|_1 \qquad \text{subject to } Y_1 = \alpha_1 X_1 \\ &\min \|\alpha_2\|_1 \qquad \text{subject to } Y_2 = \alpha_2 X_2 \end{aligned} \tag{15}$$

where $\|\,.\,\|_1$ represents the L_1 norm. Since the projections α_1, α_2 may be of very high dimensionality, it is assumed that

$$\min \|\alpha_1\|_1 = |\alpha_{1_1}| + |\alpha_{1_2}| + \cdots |\alpha_{1_n}| \tag{16}$$

Therefore the L_1 penalty is

$$\frac{\partial \min \|\alpha_1\|_1}{\partial Y_1} \tag{17}$$

further

$$\begin{aligned} \tfrac{\partial|\alpha_1|}{\partial Y_{1_1}} &= \textstyle\sum_{i=1}^{n} \tfrac{\partial|\alpha_{1_i}|}{\partial Y_{1_1}} = \sum |X_1^{-1}|_{row1} \;\; \text{sign}|Y_{1_1}| \\ &\vdots \\ \tfrac{\partial|\alpha_1|}{\partial Y_{1_i}} &= \textstyle\sum_{i=1}^{n} \tfrac{\partial|\alpha_{1_i}|}{\partial Y_{1_i}} = \sum |X_1^{-1}|_{rowi} \;\; \text{sign}|Y_{1_i}| \end{aligned}$$

resulting in

$$\frac{\partial \min \|\alpha_1\|_1}{\partial Y_1} = \sum |X_1^{-1}| \;\; \text{sign}|Y_1| \tag{18}$$

C. Stopping Criteria

All iterative optimization methods require stopping criteria to indicate the successful completion of the process. Consider,

$$\delta = \frac{max(\Delta_{NN}) - min(\Delta_{NN})}{max(\Delta)} \tag{19}$$

where, the term Δ_{NN} is the nearest neighbor distance in the resulting output distribution, Δ is the distance between any two samples in the output distribution, max(.) and min(.) are the maximum distance and minimum distance between samples in the output space. The numerator is a measure of uniformity of the output space and the denominator is a measure of how well the output space is filled. Therefore, (19) can be used as a convergence criterion. However, Δ is dependent on the number of samples obtained from the signal n, the dimensionality N and the size of the output space d. As the numerator approaches zero for uniformly

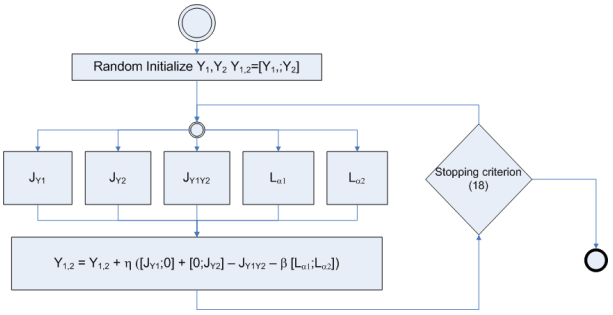


Fig. 1. Block Diagram of Proposed Method, η is the learning rate, β is the normaliser on the L_1/L_2 penalties applied to the projection coefficients α_1 and α_2.

distributed samples and for a given threshold γ required, δ may be determined by $\frac{\gamma}{d^{-N}/n}$. For all experiments in this paper following parameter values have been chosen.

TABLE I
OPTIMIZATION LEARNING RATE COEFFICIENTS

η	$\frac{M_1 M_2}{N_1 N_2}$
β	$\max\left(max(X_2)\frac{N_2}{N_1}, max(X_1)\frac{N_1}{N_2}\right)$

IV. SIMULATION AND EXPERIMENTAL RESULTS

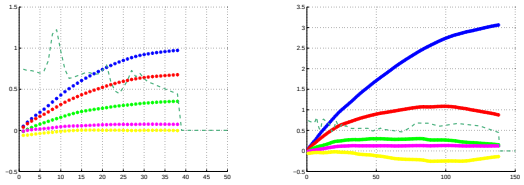
For the simulation and experimental study, output space dimensionality is chosen to be $d = 2$. For a sample size, $n = 100$, the stopping criteria from equation (19) is calculated to be $\delta < 0.035$. In order to detect that the optimization has reached a local minima the variation of δ should be contained in a $1.5e^{-3}$ limit at least for a minimal convergence span of 5 iterations.

A. Simulation Results

Two simulations are performed to evaluate the proposed method. Simulation 1: The purpose is to detect identical signal pairings embedded within a number of unrelated signals. Simulation 2: The purpose is to identify non informative signals. We have utilised Johnson's [12] method of generating signals with an arbitrary high order of dependency. Signals that are generated for the purpose of simulation are scaled to $[-1, 1]$.

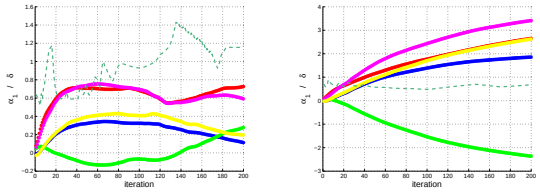
Simulation 1: Identical Signals: One hundred signals are generated, containing 100 samples each. Five signals are selected and supplied as sensor 1 output $\{1, 2, 3, 4, 5\}$ and one signal is selected as sensor 2 output $\{1\}$, thus, $N_1 = 5$ and $N_2 = 1$ with one signal in common.

In order to determine the most informative signal we examine the vector of α_1 co-efficients, where each α_{1_i} corresponds to a X_{1_i}. Results are presented in Fig. 2 with the mapping coefficients, α_{1_i} $i \in \{1, 5\}$ in blue, red, green, cyan and yellow respectively. The convergence criterion, δ is plotted as the *dashed* gray line. The results show the highest coefficient for α_{1_1} confirming that signal 1 is common between the sensors. Applying the L_1 norm penalty



(a) L1 penalty (b) L2 penalty

Fig. 2. Results of indirect estimation of Mutual Information for signals with underlying linear dependency



(a) L1 penalty (b) L2 penalty

Fig. 3. Results of indirect estimation of MI for non informative signals

to the optimization produced faster convergence, occurring in iteration 38 compared to 142 iteration with L_2 norm penalty. It is to be noted that only the non-zero mapping parameter ideally should be α_{1_1} and all others should be zero. However, due to the approximations in the objective function and the presence of local minima, the other mapping parameters have smaller non-zero vales.

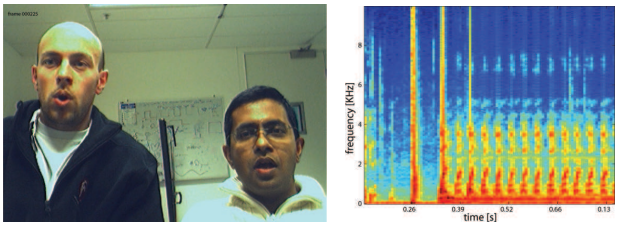
Simulation 2: Non Informative Signals

In this simulation signals $\{1, 2, 3, 4, 5\}$ are selected as sensor 1 output and signal $\{6\}$ is chosen as the sensor 2 output, clearly there is no common signals. Fig. 3 shows that neither L_1 or L_2 norm penalty has produced convergence in 200 iterations. In fact the solution based on the L_1 regularization shows a divergence from an optimized solution verifying there is no common signal.

B. Experiments

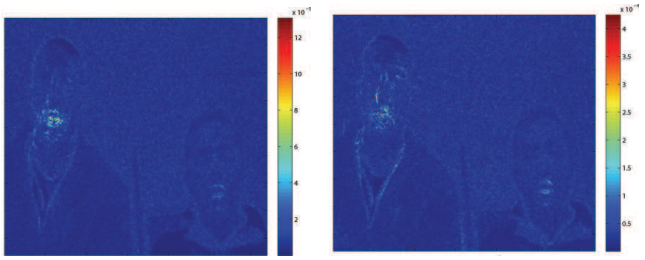
Two experiments are performed to evaluate the proposed method in establishing the relationship between multi-modal sensory data by identifying informative signals without any prior knowledge about geometric parameters. Experiment 1: The purpose is to localise the audio source in the video data sequence. Experiment 2: The purpose is to identify the common source in a laser and video data stream.

Experiment 1: Audio and Video Signals: A microphone and camera were used to capture activity in an office environment consisting of a person (left on image) reading a sequence of numbers and another person (right of image) mimicking unscripted sentences (see Fig. 4(a)). Video data was captured at 15Hz while audio signal was captured at 48KHz with only 10KHz of content used. Both video and audio data streams were synchronised in time. Color images acquired were transformed to grey scale and pixel intensity values (consisting of $640 * 480 = 307200$ pixels per frame) of 100 frames were analyzed using raw pixel values. The



(a) (b)

Fig. 4. Samples of signal sequence (a) camera data (b) audio periodogram



(a) (b)

Fig. 5. MI projection coefficients achieved on signal level between Audio and Video using indirect MI estimation with (a) L_1 norm penalty and (b) L_2 norm penalty

audio data was transformed to a series of periodograms as shown in Fig. 4(b). The window length of the periodogram is 2/15s (corresponding to two video frames). The scenario here requires finding the most mutually informative pixels from 307200 signals from the camera to 200 signals from the audio data.

The results of application of L1 (Fig. 5(a)) and L_2 (Fig. 5(b)) regularizations show images of obtained projection coefficients where the highest values denote areas of the image containing the mouth of the person sitting on the left (which is true). It is observed that the applying the L1 norm penalty to the optimization produced faster convergence, occurring in iteration 59 compared to 141 iteration with L2 norm penalty (Fig. 6)

Experiment 2: Laser and Camera Signals: A SICK laser range finder with a 180°field of view (FOV) and a camera with a horizontal FOV of 60°were used to capture motions in an office environment. Ordinary office activity consisted

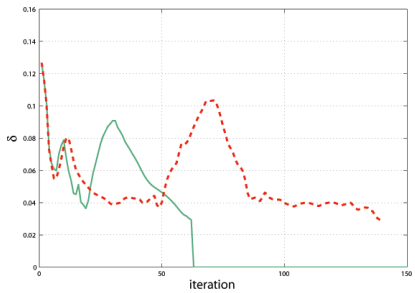
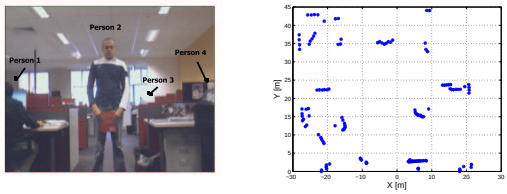
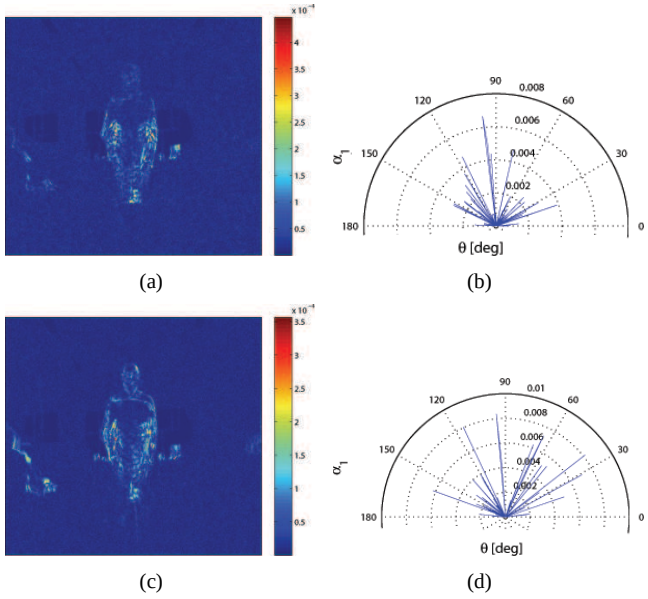


Fig. 6. Analysis of convergence properties of the L_1 (solid line) and L_2 norm penalty (dotted line)



(a) (b)

Fig. 7. Samples of signal sequence observed from (a) camera and (b) laser



(a) (b) (c) (d)

Fig. 8. Mi projection coefficients achieved on signal level localization between Laser and Camera using indirect MI estimation with (a,b) L_1 norm penalty and (c,d) L_2 norm penalty

of person 1 operating a computer mouse, person 3 moving in an office and person 4 moving at his desk Fig. 7(a). In addition, significant motions of person 2 shaking a book up and down were introduced. The laser range finder and the camera were capturing data at 75Hz and 10Hz respectively. The laser beam of the range finder intersects horizontally at the abdominal area of the standing person capturing the movement of the book.

The color images acquired were transformed to grey scale and pixel intensity values (consisting of $640 * 480 = 307200$ pixels per frame) of 80 frames were registered against 80 time synchronised raw laser readings. The scenario here requires finding the most mutually informative signals from 307200 signals from the camera to 181 signals from the laser range finder.

As discussed previously, the highest projection coefficients α_1, α_2 denote areas of highest mutual information. The results of the application of L_1 (Fig. 8(b)) and L_2 (Fig. 8(a)) regularizations show an image of obtained projection coefficients α_1, where the highest value denotes areas of the image containing the moving hands of person 2. Smaller values highlight the left most sitting person's (person 1)

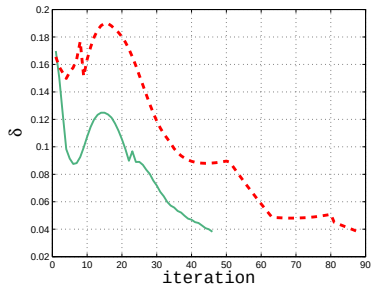


Fig. 9. Analysis of convergence properties of the L_1 (solid line) and L_2 norm penalty (dotted line)

hand and his chin movement. Fig. 8(d) and Fig. 8(d) show the projection coefficients of the laser scan α_2. There the significant peak is due to the hands of person 2. Although there seems to be some correlations with the motions present in the environment, the method correctly matches person 2 in the image sequence with person 2 in the laser sequence.

Applying the L_1 norm penalty to the optimization produced faster convergence, occurring in iteration 72 compared to 110 iteration with L_2 norm penalty. Furthermore, the coefficients α_1, α_2 have fewer non zeros, thus achieves better separation of the informative signals from other noise.

To evaluate the results of the proposed indirect MI estimation incorporating the L_1 norm penalty the experiment was repeated on 18.5 seconds worth of video data iteratively using 80 video frames and the corresponding laser returns producing 27 correct matches and 13 incorrect matches. The results reveal that registration could be performed without artificially augmenting the environment due to natural occurring movements such as person 1 moving a computer mouse or the torso and head of person 3 moving in the office cubicle. However, in some cases, changes of pixel intensity may not be directly linked to the change of range to the object unless the experiment is performed in an environment where luminance is altered with the distance to the object. This had an influence on the 13 incorrect matches. Alternatively raw data can be processed and feature level signals can be used to improve the registration results [4].

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we have formulated the sensor registration problem with a method that detects the sensor signal pairing via indirect estimation of mutual information. As opposed to the L_2 regularisation, which commonly used in the literature, we have introduced a L_1 regularization which concentrates energy to few dominant components rather than spreading over many valued components. This leads to faster convergence with less spurious correlations when compared to the use of L_2 regularization. Experiments and simulations were carried out to validate the findings.

Research in several directions to extend the work presented in this paper are currently under way. Formulating the problem in the feature level rather than signal level will remove the requirement of preserving locality of the data source. Combining the indirect estimation methods with direct estimation could couple their respective strengths and would be a fruitful avenue of further research into signal grouping. Constructing a multidimensional feature space by combining the separate features could add value and this would obviously benefit future research outcomes.

VI. ACKNOWLEDGMENTS

This work is supported by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government. Gamini Dissanayake wishes to acknowledge Stephen Boyd for providing a deep insight into the properties of L_1 regularization.

REFERENCES

[1] M. Aboelaze and F. Aloul. Current and future trends in sensor networks: a survey. *Wireless and Optical Communications Networks, 2005. WOCN 2005. Second IFIP International Conference on*, pages 551–555, March 2005.

[2] A. Alempijevic, S. Kodagoda, and G. Dissanayake. *Sensor Registration for Robotic Applications*, pages 233–242. Springer Berlin / Heidelberg, 2008.

[3] A. W. Bowman and A. Azzalini. *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, volume 18 of *Oxford statistical science series*. Oxford University Press, Oxford OX2 6DP, UK, 1997.

[4] T. Butz and J. P. Thiran. From error probability to information theoretic (multi-modal) signal processing. *Signal Process.*, 85(5):875–902, 2005.

[5] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2008.

[6] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 1991.

[7] J. Fisher and T. Darrell. Speaker association with signal-level audiovisual fusion. *Multimedia, IEEE Transactions on*, 6(3):406–413, June 2004.

[8] J. Fisher and J. Principe. Entropy manipulation of arbitrary nonlinear mappings, 1997. citeseer.ist.psu.edu/fisher97entropy.html.

[9] E. Hansen and G. Walster. Solving overdetermined systems of interval linear equations. *Reliable Computing*, 12(3):239–243, June 2006.

[10] D. Hoyer, B. Pompe, K. Chon, H. Hardraht, C. Wicher, and U. Zwiener. Mutual information function assesses autonomic information flow of heart rate dynamics at different time scales. *Biomedical Engineering, IEEE Transactions on*, 52(4):584–592, April 2005.

[11] A. T. Ihler, J. W. Fisher, and A. S. Willsky. *Hypothesis Testing over Factorizations for Data Association*. Information Processing in Sensor Networks. Springer, New York, 2003.

[12] E. W. Johnson. *Forest Sampling Desk Reference*. CRC Press, 2000.

[13] W. Landman and E. Klopper. 15-year simulation of the december to march rainfall season of the 1980s and the 1990s using canonical correlation analysis (cca). *Water SA*, 24:281–285, 1998.

[14] N. A. Nabih. L1 solution of overdetermined systems of linear equations. *ACM Trans. Math. Softw.*, 6(2):220–227, 1980.

[15] L. Navarro and E. M. Quilis. Exploring the spanish interbank yield curve. Working Papers 25-03 Classification-JEL, Instituto de Estudios Fiscales, 2003.

[16] J. Pluim, J. Maintz, and M. Viergever. Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on*, 22(8):986–1004, Aug. 2003.