

© [2009] IEEE. Reprinted, with permission, from [Alempijevic, A.; Kodagoda, S.; Dissanayake, G. Mutual information based data association. Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2009 5th International Conference]. This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Technology, Sydney's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it

Mutual Information Based Data Association

Alen Alempijevic, Sarath Kodagoda, Gamini Dissanayake

*ARC Centre of Excellence for Autonomous Systems, University of Technology Sydney
Sydney, Australia*

a.alempijevic@cas.edu.au, s.kodagoda@cas.edu.au, g.dissanayake@cas.edu.au

Abstract—Relating information originating from disparate sensors without any attempt to model the environment or the behaviour of any particular object within it is a challenging task. Inspired by human perception, the focus of this paper will be on observing objects moving in space using sensors that operate based on different physical principles and the fact that motion has in principle, greater power to specify properties of an object than purely spatial information captured as a single observation in time. The contribution of this paper include the development of a novel strategy for detecting a set of signals that are statistically dependent and correspond to each other related by a *common cause*. Mutual Information is proposed as a measure of statistical dependence. The algorithm is evaluated through simulations and three application domains, which includes, (1.) Grouping problem in images, (2.) Data association problem in moving observers with dynamic targets, and (3.) Multi-modal sensor fusion.

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 an integral part in most of the environments. 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. In this regard, 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 parts of a particular signal to that best correlate with another part of the other signal, acquired by a different sensor.

This type of analysis is reported in various fields including, biomedical engineering, climatology, network analysis and economy. In biomedical engineering research, heart rate fluctuations are examined against several interacting physiological mechanisms including visual cortex activity, respiratory rate etc [2] 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 [3].

There are number of techniques that could be suitable for detecting such 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. In most applications, 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 extended 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 [4], [5]. 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 [6]. However, this requires pair wise comparisons of all permutations of input signals to find the most informative statistically dependent pairings. As our focus of this paper is to solve the data association problem of small number of signals, the direct estimation of MI is a feasible solution. In this paper, we have presented three applications. In the first application, image features that are originated from a common cause are analyzed to solve the feature grouping problem. The second example solves track to track data association in multiple observers exploiting commonly seen targets. In the third example, fusion of multi-modal sensory data without the knowledge of geometric calibration parameters are presented.

The remainder of this paper is organised as follows. Section II discusses mutual information as a technique for detecting statistical dependence between signals. Section III outlines a method of estimating a PDF of a random variable. Solution to the integration problem is given in Section IV. Simulation and experimental results are presented in the Section V. Section VI concludes the work providing future research direction.

II. MUTUAL INFORMATION

Identification of even a single object without interaction and externally provided preconceived excitation is challenging. However, humans and animals inherently have the ability to separate objects depending on their spatial and/or temporal behaviour [7]. Objects may be perceived in accordance with similarity, continuity and overall symmetry. The principle of *common cause* particularly in visual perception, states that areas experiencing motion in the same way are connected, stimulus relations underline the ability of human perception to incorporate spacial and kinematic information in a unified framework. In line with this thinking, we utilize mutual information to localize related information resulted from a common cause in two signal streams.

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 secondly, the integration it over the set of all possible outcomes.

III. ESTIMATION OF THE PDF OF A RANDOM VARIABLE

An appropriate functional form to describe the distribution associated with the random variable is usually unknown, thus, generating a complete description of the random variable requires nonparametric estimation of the PDF. A substantial study of density estimators was conducted by Khan et. al. [8] and it was reported that kernel density estimators are the best choice for approximating a PDF of very short data at relatively high noise to signal levels. A kernel density estimate where a Gaussian kernel is adopted as a kernel function (3) is noted as parzen density estimator (PDE) [9]. The Gaussian kernel selection results in a continuously differentiable estimator which is of particular use for gradient descent based optimization [8].

$$p(x) = \frac{1}{N} \sum_{i=1}^N G(x - x_i, \Sigma) \quad (3)$$

$$G(x, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} \Sigma^{\frac{1}{2}}} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right) \quad (4)$$

where, G is a d -dimensional Gaussian kernel (4) having a diagonal isotropic covariance matrix, $\Sigma = \sigma^2 I$ and N is the number of samples. The kernel bandwidth σ is a data

dependent heuristic [10], [11] leading to a data pre-processing step. Bowman et. al. [10] derived a robust and computationally efficient method that accommodates long tailed distributions and possible outliers by applying optimal smoothing as given by

$$\sigma = \left(\frac{4}{(d+2)n}\right)^{\frac{1}{d+4}} \frac{\text{median}(x_i - \tilde{x})}{0.6745} \quad (5)$$

where n is the number of samples, \tilde{x} is the median of these samples and d is the number of dimensions. It could be noted that the value of σ is estimated for the entire sample size and will need to be calculated for each feature vector X . Despite being robust, the PDE combined with Shannon's definition of entropy (2) for estimation of MI is still unable to provide a closed form solution [12]. The integral in Shannon's entropy (2) still needs to be numerically evaluated leading to a substantial computational cost at finite accuracy.

IV. SOLUTION TO THE INTEGRATION PROBLEM

Combining the nonparametric density estimator with an approximation of theoretical entropy has been widely described in the literature to overcome this problem and is referred to as *plug-in* estimation. Some approximation to entropy impose requirements on the random vector X or the underlying PDF. Combining the PDE with Renyi's entropy estimation does not impose any restrictions and is defined as

$$H(X) = -\frac{1}{1-\alpha} \log \int_{\Omega} p(x)^{\alpha} dx, \quad \alpha > 0, \alpha \neq 1 \quad (6)$$

A number of researchers have conducted a systematic study of the effect of Renyi's entropy order, α and suggested that Renyi's quadratic entropy (where $\alpha = 2$) is preferred over Shannon's entropy. It provides nearly identical values to Shannon's entropy values with exponentially reduced computational complexity [13]. Now combining Renyi's quadratic entropy with the PDE

$$p(x) = \frac{1}{N} \sum_{i=1}^N G(x - x_i, \sigma) \quad (7)$$

results in a new entropy estimator

$$H(X) = -\log \int_{\Omega} \frac{1}{N} \sum_{i=1}^N G(x - x_i, \sigma) \frac{1}{N} \sum_{j=1}^N G(x - x_j, \sigma) dx$$

For two Gaussian kernels,

$$\int_y G(y - a_i, \sigma_1) G(y - a_j, \sigma_2) dy = G(a_i - a_j, \sigma_1 + \sigma_2)$$

Yielding,

$$H(X) = -\log \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G(x_i - x_j, 2\sigma) \right) \quad (8)$$

The computational cost of calculating Renyi's entropy $H(X)$ is now reduced to $O(N^2)$. Hild et. al. [13] have reported that difference between Renyi quadratic and Shannon entropy exist depending on the spread of the underlying PDF. The spread is determined by analysing the kurtosis

$$k_4 = \mathbf{E}_x(x^4) - 3(\mathbf{E}_x(x^2))^2$$

Source densities can be super-Gaussian (positive kurtosis) or sub-Gaussian (negative kurtosis). Therefore, the final equation for Renyi's entropy incorporating a parzen density estimator is given by

$$H(X) = -\text{sign} \left(\sum_{i=1}^N x_i^4 - 3x_i^2 \right) \log \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N G(x_i - x_j, 2\Sigma) \right) \quad (9)$$

The additional step of calculating the sign of the kurtosis does not add to the computational cost of which is $O(N^2)$. Now, the equations for mutual information (1) and (9) can be directly used to calculate MI between two sets of signals X_1 and X_2 by pairing all signals from one sensor with each of the signals from the other. If the dimensionality of the sensory signals are very high, then the signals can be pre-processed to extract a limited number of features for pairwise comparisons.

V. SIMULATION AND EXPERIMENTAL RESULTS

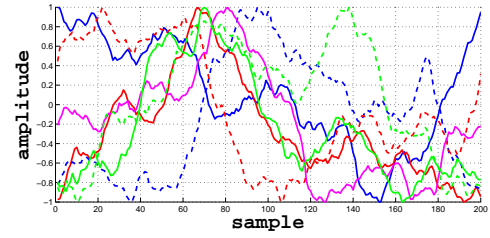
In this section the simulation results are presented as a proof of concept and the experimental results validate the algorithms in practical applications.

A. Simulation Results

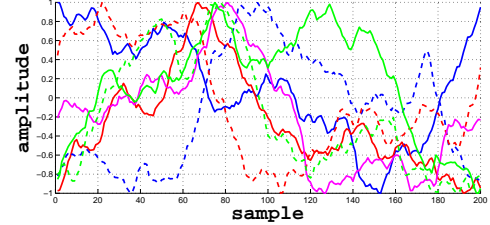
Simulations are carried out to evaluate the use of mutual information for detecting statistical dependence between artificially generated signals. Two simulation studies are conducted in order to evaluate the ability of MI to detect (1.) Identical signals among two sets of signal streams (2.) Signals that are related though a quadratic function form among two sets of signal streams.

Simulations are performed on two signal sets, each set containing a number of signals N_1 and N_2 respectively. Calculation of mutual information between the two sets is performed by pairing each signal from the two sets, therefore a total of $N_1 N_2$ pairings are constructed. The signals are generated with an arbitrary high order of dependence through evaluations of the exponential function [14]. All generated signals are scaled to $[-1, 1]$ with zero mean.

Simulation 1: Identical Signals: One hundred signals are generated, containing 8000 samples each. A set of seven signals are arbitrarily selected as output of sensor 1 $\{1, 2, 3, 4, 5, 8, 9\}$ (Fig. 1(a)) and another set of seven signals are selected as the output of sensor 2 $\{1, 2, 3, 4, 5, 6, 7\}$ (Fig. 1(b)) with five signals in common and the others are completely unrelated.

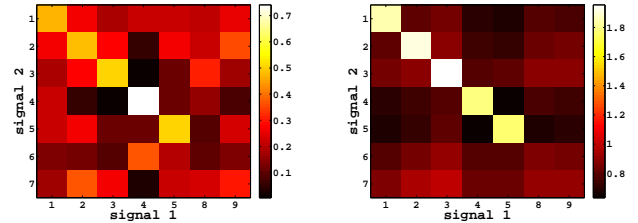


(a) Set of signals of Sensor 1



(b) Set of signals of Sensor 2

Fig. 1. Signals of sensor 1 and 2



(a) 200 samples

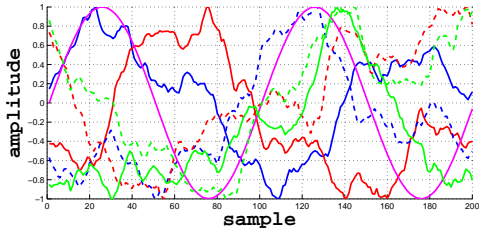
(b) 8000 samples

Fig. 2. Results of pairwise calculation of mutual information

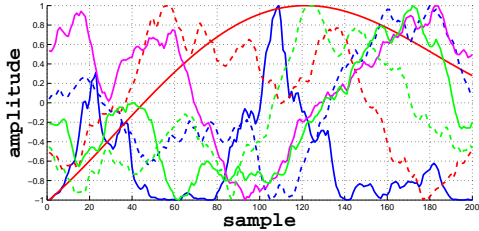
A total of 49 signal pairings are constructed and MI are calculated. The results are presented in a matrix form in Fig. 2(a). The axes on the graphs representing the matrix denote the examined signals, while each cell in the matrix contains the value of mutual information calculated for the respective pairing, colors of the matrix cells are based on the scale adjacent to the figures.

Observing the results it could be noted that the high MI values lie on the diagonal of the matrix spanning the first five rows in Fig. 2(a). It is also noted that as the size of the available data set increases (200 samples in Fig. 2(a) and 8000 samples in Fig. 2(b)), the estimated MI value monotonically grow toward the true value of mutual information. This shows that Renyi's quadratic entropy through an approximation is good enough to identify common signals. However, the computational complexity is an issue as it required 157.02 seconds to complete the MI calculation for all signal pairings with 8000 samples, although reducing to 12.1 seconds with 200 samples.

Simulation 2: Signals Related by a Quadratic Form: The objective of this study is to establish the MI's capability to detect signals even if they are not directly related, but the signals are quadratically related. We use the quadratic function given below for generating additional signals.



(a) Set of signals of Sensor 1



(b) Set of signals of Sensor 2

Fig. 3. Signals of sensor 1 and 2 with underlying quadratic dependence

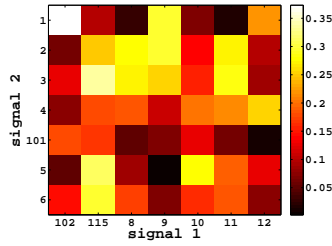


Fig. 4. Results of pairwise calculation of Mutual Information for signals with underlying quadratic dependence

$$f(x)^2 - 5x^2 = 5 \rightarrow f(x) = \sqrt{5 * (x^2 + 1)} \quad (10)$$

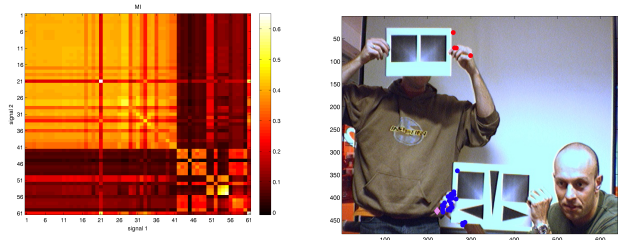
Signal 102 is derived from signal 1 using (10). While keeping signal 102 and 1 as sensor 1 and sensor 2 outputs respectively, we have added some unrelated signals to sensor 1 ($\{102, 115, 8, 9, 10, 11, 12\}$) and sensor 2 ($\{1, 2, 3, 4, 101, 5, 6\}$). Mutual information is calculated on signal pairings with 200 samples in each. The supplied signal sets are depicted in Fig. 3.

Although it is intuitively hard to identify the signals of quadratic dependence, the results given in Fig. 4 shows that high MI values are corresponding to signal 1 of sensor 2 and signal 102 of sensor 1 validating the quadratic dependency.

B. Experiments

In this section three case studies are presented to prove the practical applicability of the algorithm. Those are, (1.) Grouping problem in images, (2.) Data association problem in moving observers with dynamic targets, and (3.) Multi-modal sensor fusion.

Experiment 1: Grouping problem in images: In this experiment, we use the MI based algorithm to group the



(a) MI matrix

(b) Grouping result

Fig. 5. Grouping of targets achieved using Mutual Information matrix for clustering

features attributed to a single body. A camera with 31° and 41° of horizontal and vertical field of view is used to capture few motions in a lab environment. Two identical boards with patterns (to introduce ambiguity) are moved differently in order to generate a rich set of features that can easily be tracked (see Fig. 5(b)). From the wide range of computer vision based feature tracking algorithms the Kanade Lucas Tomasi (KLT) tracker [15] is selected as it is fast and robust. A maximum number of 60 features are tracked. Attributes of features (bearing and elevation to the features) are calculated and used for pairwise MI calculation. The image pixels themselves can not be used as signals in the algorithm as they do not represent the motions related to the grouping problem.

The results of the pairwise calculation of MI are presented in Fig. 5(a). The axes on the graph denote the examined signals, while each cell in the matrix contains the value of mutual information calculated for the respective pairing. Signal pairings with an underlying relationship result in a similar MI value. Therefore, the MI values can be used as an inter distance metric in determining clusters of signals that are mutually dependent, which are arising from a common cause. Grouped targets are presented in Fig. 5(b), where we can observe a clear distinction of two groups of features noted with blue/red colors. The target noted with a purple diamond is an invalid track and did not belong to either target group.

Experiment 2: Data association problem in moving observers with dynamic targets: The objective of this experiment is to evaluate the use of MI for detecting dynamic targets with moving observers. Experiments were conducted in a maze like environment with five robots (Fig. 6). The maze walls are high enough to prevent the lasers seeing the targets on the other side of the wall, creating occlusions. Two SICK laser range finders were mounted on two Pioneer platforms acting as observers (O1 and O2 in Fig. 6). Reflective poles mounted on iRobot Create platforms were used as moving targets (T1, T2 and T3 in Fig. 6). All the Pioneers and the Create robots were independently controlled. The structure of the environment is assumed to be completely unknown. Two small boxes were also placed in the environment to create additional occlusions. The synchronised Pioneer platforms collected laser range data at 7.5Hz and odometry at 10Hz. Although the moving platforms can localise in their individual coordinate systems, the

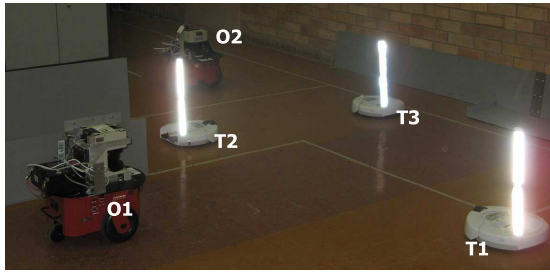


Fig. 6. Experimental setup with two Pioneers and three iRobot Creates

relationship between the two coordinate systems is completely unknown. We use an Iterative Multiple Model algorithm [4] for target tracking. The goal is to use mutual information to associate common tracks seen by both observers.

The experiment in Fig. 7 starts with one common target T2. Targets T3 is only visible to Observer O1 and target T1 is only visible to observer O2. As the observers and targets move, the target T3 becomes a second common target while target T2 becomes occluded for Observer O1 as in Fig. 7(a). While the target is occluded, the IMM estimates the target pose. Fig. 7(a) and Fig. 7(b) show the observer localization history and tracked target history in individual coordinate frames. Attributes of the targets, such as range and bearing to the target are estimated to be used in the MI algorithm. The MI based algorithm could establish the relationship with attributes of some of the targets commonly seen by the two observers. Once a common target is known by the two observers, the trajectories can be used for estimating the relative localizations of the observers. This can be verified by transforming data of observer O2 on to the co-ordinate frame of observer O1 and vice versa as shown in Fig. 7(c) and Fig. 7(d). It could be seen that the transformed range-bearing information is aligned with the observed data confirming correct target matching. This enables sharing of the entire track history of all targets between observers enhancing field of view of each observer. For example, as in Fig. 7(d) Observer O2 is now aware of the entire trajectory of Target T3 including travel prior to entering its field of view.

Experiment 3: Multi-modal sensor fusion: The aim of the experiment 3 is to evaluate the MI based algorithm for multi-sensor data fusion applications. 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 of person 1 operating a computer mouse, person 3 moving in an office and person 4 moving at his desk (see Fig. 8(a)). In addition, significant motions of person 2 shaking a book up and down were introduced. Samples of the signal sequence are presented in Fig. 8. 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 aim is to localize the most mutually informative parts of the signals based on the dynamics in the environment. No information

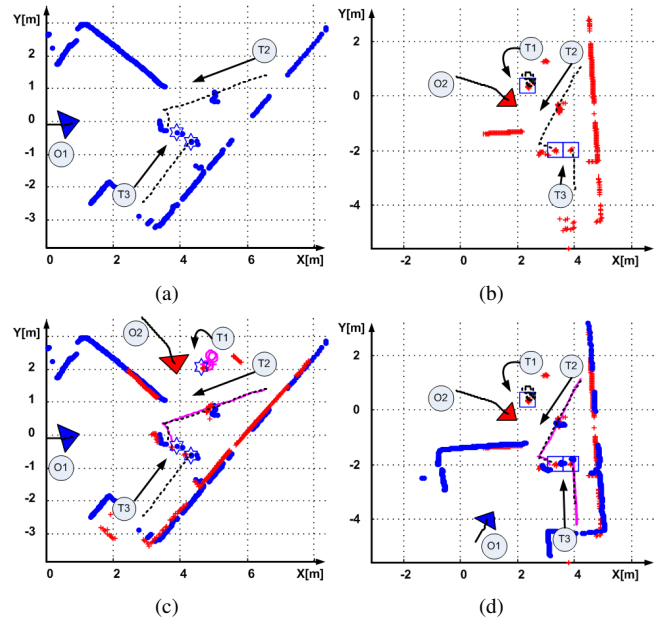


Fig. 7. Sensor registration with moving observers and moving targets experiencing occlusions (a) Test environment: Observer O_1 world coordinate frame, (b) Test environment: Observer O_2 world coordinate frame, (c) All data transformed into Observer O_1 reference frame, O - observer, T - target and (d) All data transformed into Observer O_2 reference frame

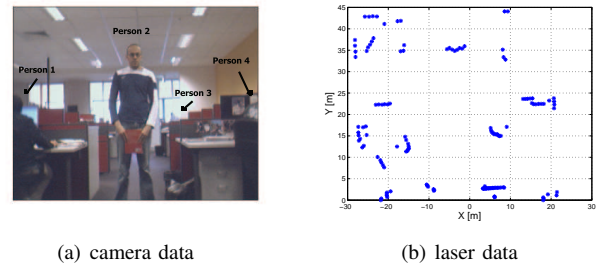


Fig. 8. Samples of signal sequence observed

about the relative sensor locations were assumed.

Pixel intensity values of grey scale images acquired (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. Grey scale images are initially transformed into column vectors of 307200 pixel values. Vectors from successive time instances are collated together to form a matrix. In this manner, successive observations from each pixel of the camera were represented as signals in each row of the set of signals X_1 , thus X_1 is interpreted as a matrix consists of 307200 signals. The set of signals X_2 originating from the observations from the laser range finder were constructed in a similar fashion.

The pairwise approach is applied to the two sets of signals: pixel variation sequences from the set of signals X_1 and range variations from the set of signals X_2 creating a total of $5.57 \cdot$

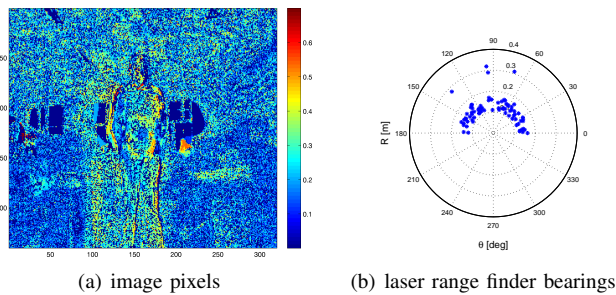


Fig. 9. Multi-modal sensor fusion with MI values related to each respective sensor

10^7 comparisons.

The results of the experiment are presented as follows. MI values related to the camera are represented as an image where larger coefficients denote areas of highest MI as in Fig. 9(a). Similarly, MI values obtained for the laser range finder are represented with values at each bearing reading as in Fig. 9(b).

From the results of direct MI estimation shown in Fig. 9, it can be observed that the highest MI values denote areas of the image containing the moving hands of person 2. Smaller values highlight the left most sitting person's (person 1) hand and chin movement. Fig. 9(b) shows the MI values of the laser scan, where significant peaks are due to the hands of person 2, while the chin and hand of person 1 produce smaller peaks. Therefore, by intuitively comparing values of highest MI, the method correctly matches person 2 in the image sequence with person 2 in the laser sequence producing the results Fig. 9. Therefore, this is a very strong application of MI, where multi-modal sensory data can be fused to detect mutually informative signals even without knowing any geometrical relationships of sensors. However, downside of this particular application, where large number of signals are involved is that the computational complexity (it required 109 hours). Accuracy of the algorithm was also compromised by complex relationship of the laser range data and corresponding intensity variations. However, both the computational complexity and accuracy can be improved by introducing feature level signals, such as optical flows associated with pixels and tracked targets using laser range data.

VI. CONCLUSIONS AND FUTURE WORKS

This paper presented a direct method of estimation of mutual information between sensory signals, which is used to identify mutually informative signals. The concept has been validated through simulations and three practical applications. Firstly, MI was used to categorize image features, which are belonging to the same object (grouping problem) based on their dynamics. In the second application, MI was used to resolve the data association problem in moving observers in a dynamic environment. Third application solved the multi-modal sensor fusion problem without having to geometrically relate the sensors. All the applications showed the feasibility of using dynamics present in the environment along with the MI based algorithm for data association.

As the direct MI based algorithm requires the sensory signals to be pairwise compared, this method is not feasible for situation where large number of signals are present. In such applications, it is advisable to extract the features (and hence lowering the number of signals) and carry out the pairwise comparisons. As future work, we are planning to extend the study to other domains of applications, such as correlating physiological data with camera and laser range finder data to analyze driver fatigue status.

ACKNOWLEDGMENT

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.

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*, pp. 551–555, March 2005.
- [2] 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*, vol. 52, no. 4, pp. 584–592, April 2005.
- [3] 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*, vol. 24, pp. 281–285, 1998.
- [4] A. Alempijevic, S. Kodagoda, and G. Dissanayake, *Sensor Registration for Robotic Applications*. Springer Berlin / Heidelberg, 2008, pp. 233–242.
- [5] J. Fisher and T. Darrell, "Speaker association with signal-level audio-visual fusion," *Multimedia, IEEE Transactions on*, vol. 6, no. 3, pp. 406–413, June 2004.
- [6] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
- [7] P. Kellman and T. Shipley, "A theory of visual interpolation in object perception," *Cognitive Psychology*, vol. 23, pp. 141–221, 1991.
- [8] S. Khan, S. Bandyopadhyay, A. R. Ganguly, S. Saigal, D. J. Erickson, III, V. Protopopescu, and G. Ostrouchov, "Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data," *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 76, no. 2, p. 026209, 2007.
- [9] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 1991.
- [10] A. W. Bowman and A. Azzalini, *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, ser. Oxford statistical science series. Oxford OX2 6DP, UK: Oxford University Press, 1997, vol. 18.
- [11] E. Parzen, "On the estimation of a probability density function and mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.
- [12] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symposium Mathematics Statistics and Probability*, vol. 1. University of California Press, Berkeley, CA, 1961, pp. 547–561.
- [13] K. E. Hild, D. Erdogmus, and J. C. Principe, "An analysis of entropy estimators for blind source separation," *Signal Process.*, vol. 86, no. 1, pp. 182–194, 2006.
- [14] E. W. Johnson, *Forest Sampling Desk Reference*. CRC Press, 2000.
- [15] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 1981, pp. 674–679.