Mutual Information based Sensor Registration and Calibration

Alen Alempijevic[*], Sarath Kodagoda[*], James Underwood[+], Suresh Kumar[+], Gamini Dissanayake[*]

ARC Centre of Excellence for Autonomous Systems (CAS)

[*]Faculty of Engineering, University of Technology Sydney
NSW, Australia

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

[+]Australian Centre for Field Robotics, University of Sydney
NSW, Australia

{s.kumar, j.underwood} @cas.edu.au

Abstract **- Knowledge of calibration, that defines the location of sensors relative to each other, and registration, that relates sensor response due to the same physical phenomena, are essential in order to be able to fuse information from multiple sensors. In this paper, a Mutual Information (MI) based approach for automatic sensor registration and calibration is presented. Unsupervised learning of a nonparametric sensing model by maximizing mutual information between signal streams is used to relate information from different sensors, allowing unknown sensor registration and calibration to be determined. Experiments conducted in an office environment are used to illustrate the effectiveness of the proposed technique. Two laser sensors are used to capture people mobbing in an arbitrarily manner in the environment and MI from a number of attributes of the motion are used for relating the signal streams from the sensors. Thus the sensor registration and calibration is achieved without using artificial patterns or pre-specified motions.**

Index Terms – mutual information, sensor registration, sensor calibration

I. INTRODUCTION

The advantage of using a variety of sensors in perception and control is often nullified by the complexity in extracting and relating useful information, when the geometric relationship between sensors and underlying joint probability distribution between sensor signals are unavailable. Probabilistic data association strategies and complex calibration procedures are usually required to be able to convert information from multiple sensors to a common reference frame. Once this is achieved sensor fusion to perform a variety of tasks in robotics, sensor networks and biomedical engineering become feasible using a relatively straightforward Bayesian framework.

Mutual information (MI) analysis has well established grounds in the field of medical imaging [1] as a method for image registration and feature selection from a diverse range of sensor modalities. Further, MI has been used as a method to detect statistical relationship or a measure of coupling between signals [2]. However, traditional MI based approaches have a substantial computational cost in order of a factorial of the number of samples making those unsuitable for real time applications. Fisher et al [3] proposed a methodology for substantially reducing the computational cost of using MI. This method is based on unsupervised learning of a nonparametric sensing model by maximizing mutual information between signal streams in a lower dimensional space. Experiments where a mouth of a speaker on a video stream is pinpointed by analyzing information from video and audio signals have also been presented. Algorithm by Fisher et. al., however, requires a stationary speaker. To cope with a moving speaker Ikeda et al [4] propose a target tracker based on background segmentation, assuming piecewise constant velocity of the source and employing a heuristic search on all possible target paths. Butz [5] conducts pre-processing of images by calculating optic flow in neighboring pixels to eliminate effects of moving background objects. However, neither method deals with multiple common sources in the sensory signals. These limitations are substantial in most robotics applications, for example, fusion of data from sensors on-board of a moving vehicle with that extracted from sensors present in the road infrastructure and other near by vehicles. This paper, extends Fisher's [3] implementation to such an application where dynamic objects in the scene are utilized for sensor registration and calibration.

Moving objects in a sensory space provide very little instant information that can be easily exploited to register two signal streams. However, if these objects are tracked as moving features, a substantial amount of information in the feature level can be extracted to achieve signal registration. Attributes of features should be carefully selected giving due regards to sensor modality and rate of change of the attribute. It is important that at least one attribute of the feature has a rate of change for all possible movements of the feature in a sensory space. This allows the MI based approach to be utilized to register the common feature attributes in two sensory signals providing sensor registration.

In Section II, we examine the information theoretic approach for sensor registration. Section III describes the process of sensor calibration. Experimental results are presented in Section IV. Section V concludes the paper outlining future work.

II. Sensor Registration

In this section we describe an information theoretic approach for sensor registration, which is based on [3, 6-10]. One important aspect of this method is it does not assume any prior information about relationships among either sensory signals or sensors.

A. Mathematical Formulation

The concept of Mutual Information (MI) can be traced back to Shannon's definition of entropy (1) and can be interpreted in three possible ways: the amount of information an event gives when it takes place, the uncertainty about the outcome of an event and the dispersion of the probabilities with which the events take place.

$$H = \sum_i p_i * \log \frac{1}{p_i} \tag{1}$$

Mutual information defined in (2) can be interpreted as the amount of information *A* contains about *B* where *H(A)* is the Shannon entropy of *A* and *H(A,B)* is the joint entropy. The term *-H(A,B)* means that the mutual information can be maximized by maximizing individual entropic terms *H(A)* , *H(B)* and minimizing the joint entropy.

$$I(A,B) = H(A) + H(B) - H(A,B) \tag{2}$$

Obtaining individual entropy and joint entropy is a complex task as assumptions need to be made on the underlying PDF of the input space. To overcome these problems Fisher and Principe [3] suggest an unsupervised learning method by which multi-dimensional signals can be nonlinearly transformed onto a maximum entropy feature space resulting in statistically independent features. The method relies on an indirect measure of entropy rather than a direct estimate and exploits the principle of Information Maximization [9] which seeks to transfer maximum information about the input signal to the output features.

The goal of this approach is to choose the projection coefficients to optimize the fusion criteria, which is the mutual information [8]. Given a random vector (RV) $X \in R^N$, differential entropy, *h(x)* can be defined as in (3), where $p_x(.)$ is the probability density function of the RV. If the RV is transformed by a mapping $y=g(\alpha,x)$: α being the projection coefficient, then the entropy of new RV satisfies (4), where J_{XY} is the Jacobian of the mapping. This implies that the transformation can lead to an increase in the conveyed amount of information. The method applied seeks a subspace mapping that is in some measure optimal for classification[3].

We can estimate the density in the low-dimensional space using a nonparametric kernel based method, the Parzen window density estimator (5), where κ(.) is the kernel function and must be a valid PDF (in our case Gaussian PDF), y_i ($i=1:N_y$) is the i^{th} observation of the mapping in the output space and *u* is the location at which the output estimate is being computed.

$$h(x) = -\int_{-\infty}^{\infty} p_X(x) \log\left(p_X(x)\right) d_x \tag{3}$$

$$h(y) \le h(x) + E\left\{\ln\left(\left|J_{XY}\right|\right)\right\} \tag{4}$$

$$\hat{f}_Y(u,y) = \frac{1}{N_y} \sum_{i=1}^{N_y} \kappa(y_i - u) \tag{5}$$

We need to adapt the projection coefficient α in (6) such that the output distribution is as close to possible to uniform, hence, maximizing the entropy of the RV. As suggested by Fisher et al [3], the minimization criteria is defined as the integrated squared error between the estimated and desired distribution at a point *u* in the output space over a set of observations *y* and the uniform distribution $f_Y(u)$ as in (6).

$$\begin{aligned} J &= \int_{\Omega_Y} \left(f_Y(u) - f^{\wedge}(u,\{y\})^2 \right) du \\ J &= \frac{1}{2} \int_{\Omega_Y} \left(f_Y(u) - f^{\wedge}\left(u, g\left(\{x\},\alpha\right)\right)\right)^2 du \end{aligned} \tag{6}$$

The gradient of the optimization criteria with respect to the mapping parameters is determined via the chain rule (7), where $\varepsilon_Y(u_Y,y)$ is the computed distribution error over all observations *y* and ∂g/∂α is sensitivity of the mapping The remaining partial derivative ∂f^/∂g can be computed according to (8).

$$\begin{aligned} \frac{\partial J}{\partial \alpha} &= \left(\frac{\partial J}{\partial f^{\wedge}}\right)\left(\frac{\partial f^{\wedge}}{\partial g}\right)\left(\frac{\partial g}{\partial \alpha}\right) \\ \frac{\partial J}{\partial \alpha} &= \frac{\Delta u}{N_Y} \sum_j \left(f_Y(u_j) - f_Y^{\wedge}(u_j, y)\right)\left(\frac{\partial f^{\wedge}}{\partial g}\right)\left(\frac{\partial g}{\partial \alpha}\right) \\ \frac{\partial J}{\partial \alpha} &= \frac{\Delta u}{N_Y} \sum_j \varepsilon_Y(u_j, y)\left(\frac{\partial f^{\wedge}}{\partial g}\right)\left(\frac{\partial g}{\partial \alpha}\right) \end{aligned} \tag{7}$$

The 2nd dimensional Gaussian kernel κ' evaluated at *u* can be calculated by (9)

$$\frac{\partial f^{\wedge}}{\partial g} = \frac{1}{N_Y} \sum_{i=1}^{N_Y} \kappa'\left(g(\alpha, x_i) - u_j\right) \tag{8}$$

$$\kappa(y_i - u) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(y_i - u)^T (y_i - u)}{2\sigma^2} \right) \tag{9}$$

The adaptation amount of the mapping parameter, Δy in (10) can now be determined by the evaluated partial derivative of the kernel in (9) substituted in (8) and (7). The terms f_r and

κ_a are the attractor kernel and topology regulating term respectively.

$$\Delta y = f_r\left(y_i\right) - \sum_{i\neq j} \kappa_a\left(y_i - y_j\right) \tag{10}$$

Since the PDF estimation is now bypassed by iteratively calculating Δy (10), we can no longer evaluate the optimization criterion (6). Therefore, a measure proposed by Fisher [11] evaluates convergence by examining the projected values (y) in the lower dimensional output space as in (11). The iteration process is repeated until a local minimum is reached. The term Δ_{NN} is the nearest neighbour distance and max(Δ) is the maximum distance between any two points 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.

$$\delta = \frac{\max\left(\Delta_{NN}\right) - \min\left(\Delta_{NN}\right)}{\max\left(\Delta\right)} \tag{11}$$

B. *Adaptation to the existing algorithm*

The Fisher's [3] algorithm estimates the Mutual information between two input signals in the absence of an estimated PDF of either signal. Maximization of Mutual Information is achieved in the locally created subspace. The proposed method has been applied to signal-level fusion of audio-visual information in order to locate a speaker [10]. However, it is assumed that there is no relative motion in speaker position in the image. This assumption often violates practical applications, such as the one described in this paper, which utilizes moving objects for sensor registration. Moving objects wander across the sensory space providing very little instant information in signal level. However, they provide substantial amount of information in the feature level, if the features can separately be tracked. Therefore, we propose to use feature level registration. First the dynamic features in the sensors are extracted and attributes are estimated. The selection of the feature attributes depends on the sensor modality. An important aspect in selecting attributes is the requirement of a rate of change at least in one attribute for all possible movements of the feature. For example one may choose attributes in image features as bearing to the target, optical flow and area of occupation, whilst another may choose laser feature attributes as bearing to the target, target velocity and angle of occupation.

III. SENSOR CALIBRATION

Although the above detailed method for sensor registration can be used for many sensor modalities including camera and laser range finder, inter-sensor calibration becomes complex depending on the modality of the observation. Here we focus our attention in calibrating two horizontally mounted laser range finders.

The laser rangefinders are positioned at O_1 and O_2 observing a common point $P(X,Y), Q(x,y)$ as in Fig1, the coordinate transformation is given in (12). Let us consider there are $P:\{p_i\ (i=1,...\bar{N})\}$ number of points in the O_1 coordinate frame and $Q:\{q_i\ (i=1,..\bar{N})\}$ number of points in O_2 coordinate frame. H matrix in (13) can now be calculated with $\bar{p}$ and $\bar{q}$ being the centroids of the point sets P and Q respectively. Singular value decomposition (SVD) is then performed on H to obtain the rotation R and translation T matrix as given in (14) relating the two point sets in the Least Square sense [12] providing sensor calibration. This calibration procedure does not necessarily require two or more moving objects, rather it requires temporal multiple localizations of one moving object. This is feasible as the registration algorithm utilizes temporal information of the signals.

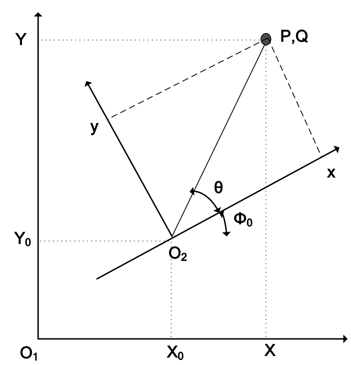


Fig. 1 Point P observed from two sensors positioned at O_1 and O_2

$$\begin{aligned} X &= X_0 + x\cos\left(\theta + \Phi_0\right) \\ Y &= Y_0 + y\sin\left(\theta + \Phi_0\right) \end{aligned} \tag{12}$$

$$\begin{aligned} H &= \sum_{i=1}^{n}\left(p_i - \bar{p}\right)\left(q_i - \bar{q}\right) \\ H &= U\Omega V^T \end{aligned} \tag{13}$$

$$R = VU^T; T = \bar{q} - R\bar{p} \tag{14}$$

IV. EXPERIMENTAL RESULTS

The MI based sensor registration and calibration were evaluated in signal level as well as feature level.

A. *Signal level sensor registration*

The algorithm was first evaluated through simulations. A total of 26 signals were generated, three randomly selected signals were designated for sensor 1 and another three for sensor 2. There was only one common signal between sensors (signal 2 in Fig. 2). Calculation of Mutual Information was performed on 100 samples with a two dimensional output space and the extent of the output distribution was 2. The convergence was evaluated imposing the following criteria on

Eq. (11). Firstly, in order to detect a local minimum the standard deviation must be contained in a 5e-3 limit for a minimum of 10 iterations. In order to determine whether the result of the iteration process has produced a valid result, the ratio of (11) must be less than 0.085. This value was derived from the extent of the output distribution and the number of samples. With the imposed criteria, a match was found in 19 iterations (Fig.3). Table 1 shows the calculated MIs with the highest MIs are denoted by symbol ‘*”. It shows signal 2 of sensor 1 is most informative with signal 2 of sensor 2, which is true.

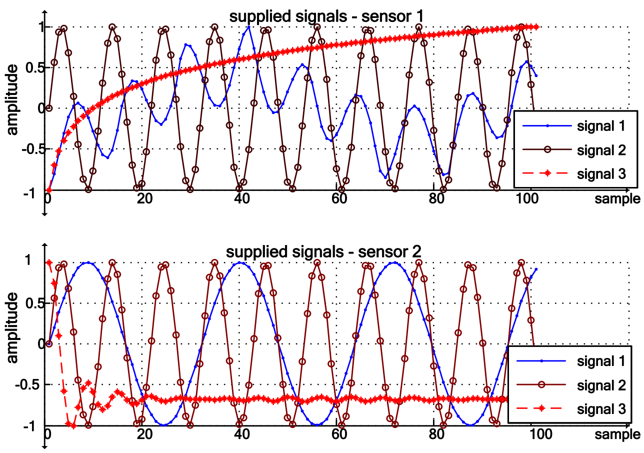


Fig 2 - Simulated signal streams

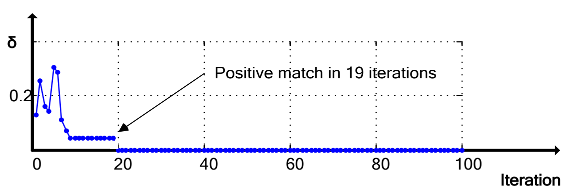


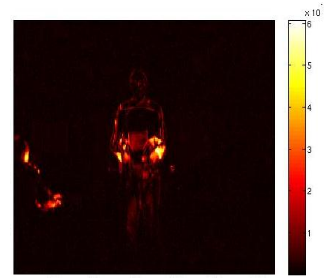
Fig 3 – Mutual Information learning curve

TABLE I

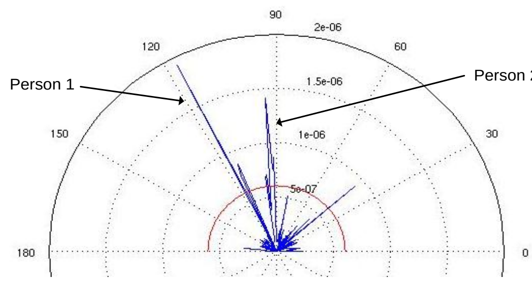
Sensor/Signal	1	2	3
1	0.0086	0.077*	0.0078
2	0.0037	0.0728*	0.004

The algorithm was also experimentally evaluated for signal level sensor registration. A laser range finder and a camera were used to capture a person shaking a book as in Fig. 4(a). The laser intersects horizontally the abdominal area of the standing person. The images acquired were transformed to grayscale, raw intensity pixel values were then fed into the algorithm as well as the raw range readings from the laser rangefinder. It is to be noted that there is some movement of the left most sitting person and another person on the right outside the field of view of the camera as well. Fig. 4(b) shows the projection coefficients of camera, which highlights the mutually informative area containing the moving hand as well as the left most sitting person’s chin movement. These areas of high mutual information lie on the intersection of the laser on the image. Fig. 4(c) shows the MI projection coefficients of the laser sensor.

(a) Single image of a sequence



(b) Projection coefficients of camera sensor



(c) Projection coefficients of laser sensor

Fig. 4 Laser/camera registration

B. Feature level sensor registration and calibration

In this section, experimental results of feature level registration and calibration are presented. Two spatially separated, horizontally mounted laser range finders were used to capture activities in an office environment, which contained at least one observable moving target. Moving objects were extracted using background subtraction as the observers were stationary. Nearest neighbour data association was used in tracking targets. Feature attributes, namely speed, angle with respect to laser coordinates and range were considered. The signal streams were fed into the MI Algorithm. The Algorithm was tested against a single moving target as shown in Fig. 5. It could be noted that the laser scan data is significantly different from each laser rangefinder due to different field of views (FOVs), which made scan matching techniques unusable. The traditional Iterative Closest Point (ICP) was unable to achieve correct registration of two laser scans due to the smaller overlap and substantial difference in FOV. Fig. 6 shows the trajectory of each target attribute with time. Feature attributes belonging to the sensors were fed into the MI algorithms one at a time and registration was achieved when δ converged for the particular feature attribute. In the experiment the speed was turned to be the most mutually informative attribute. The learning curve shown in Fig. 7 shows it finds a solution at 18th iteration. Once the moving object was registered, temporal locations of the target were fed into (11-13) for inter-sensor calibration. The transformed laser sensor 2 data onto the laser sensor 1 sensory space using the estimated calibration parameters are shown in Fig. 8. Alignment of the two laser scans in Fig 8. are a measure of correct sensor calibration.

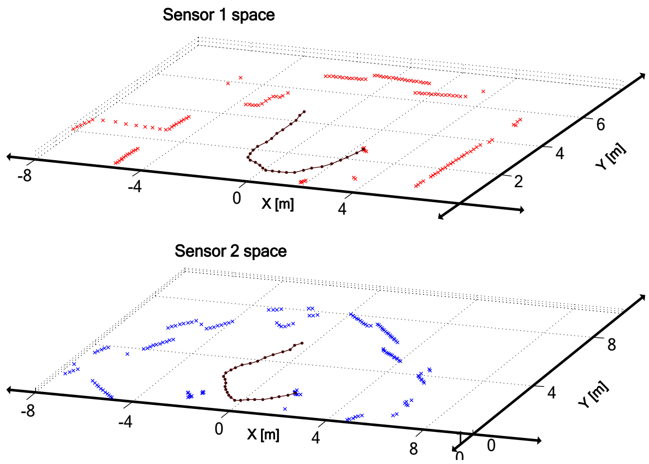


Fig 5 – Observed scene (1 moving target), *crosses* – raw laser data, *solid line* - observed trajectory

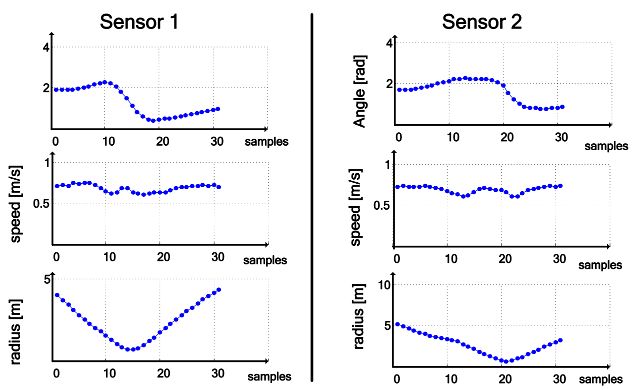


Fig 6 – Extracted Features

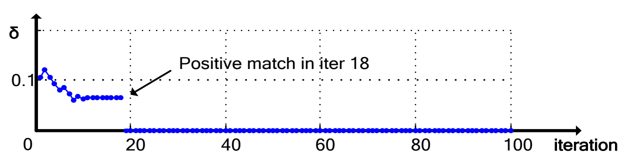


Fig 7 - Mutual Information learning curve

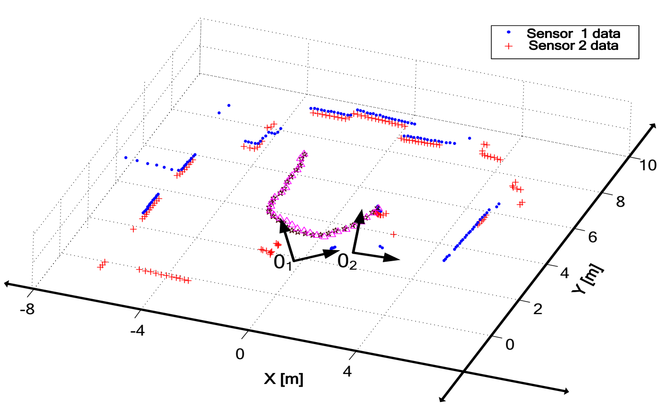


Fig 8 – Result after sensor calibration, O_1 – Origin sensor 1 and, 0_2 – Origin sensor 2

Experiment was repeated using three dynamic objects (persons) in the office environment (Fig. 9). It is to be noted that person 1 and 2 are visible to both sensors whilst person 3 is only visible to sensor 2. The attributes of the dynamic objects are shown in Fig. 10, the algorithm converged for object angle in sensory space rather than speed or range and produced high correlations. This makes the angle in sensory space to be picked as most mutually informative signals in both sensors. Further, the angle of person 1 in sensory space is most informative than that of person 2 and person 3 providing the sensor registration (see Table II). The learning curve in Fig. 11 shows that it found a match in the 20th iteration. Fig. 12 shows the aligned two sensors after sensor registration. It is to be noted that once the sensor calibration parameters are known, other two targets can be registered using the sensor geometry.

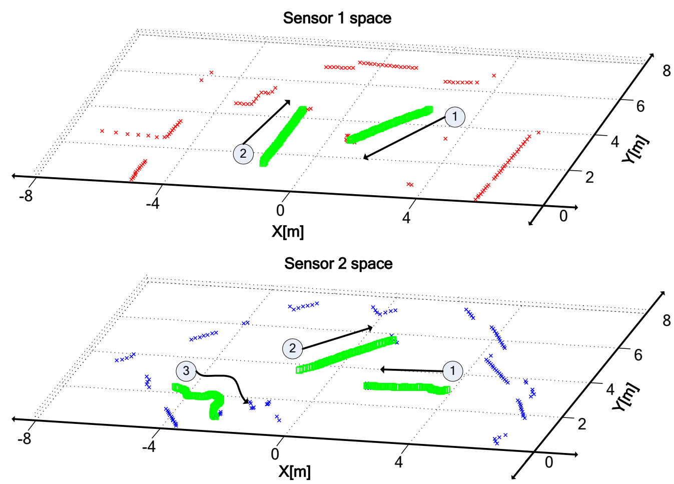


Fig 9 – Observed scene (3 targets)

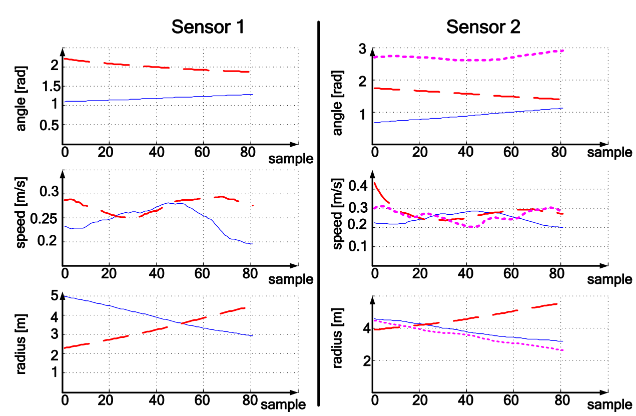


Fig 10 – Extracted Features, *solid line* – target 1, *dashed line* – target 2, *dotted line* – target 3

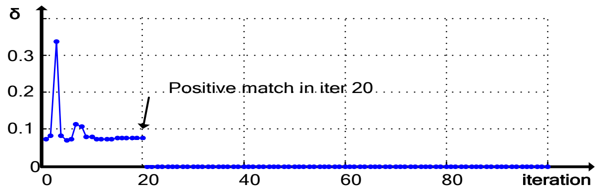


Fig 11- Mutual Information learning curve

TABLE II

Sensor \ angle	1	2	3
1	7.06*	3.6	
2	7.22*	1.02	1.29

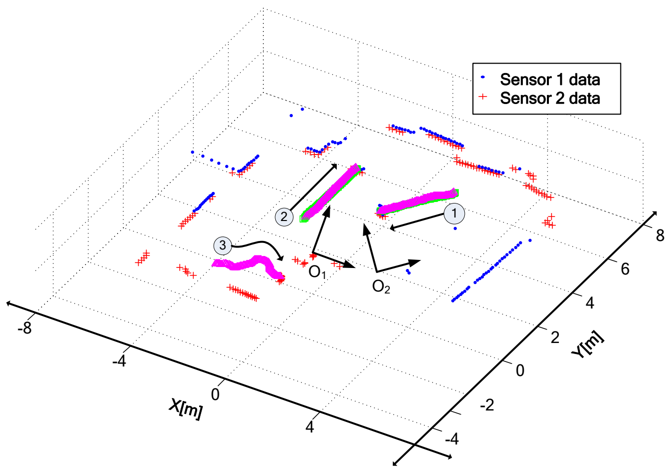


Fig 12 – Result after sensor calibration, O_1 – Origin sensor 1, O_2 – Origin sensor 2

IV. Discussion And Future Work

This paper has proposed a sensor registration method using feature level Mutual Information concepts for detecting commonalities between sensor spaces. The method utilizes a variety of attributes proprietary to detected moving objects in the examined sensory spaces to determine a relationship through an iterative method proposed by Fisher[3]. This technique does not require a specific calibration pattern to be utilized as in other sensor calibration methodologies. It only requires a moving object to be seen in both sensory spaces. Simulations and experimental results are used to validate the approach in generated signal streams as well as experimental data gathered in an office like environment using two laser rangefinders.

Further work need to be carried out in examining the convergence criteria and the underlying relationship to the number of samples and signal dimensionality. The current implementation falls short when targets experience occlusions. Further, it is not consistent with moving observers. Therefore, we are in the process of integrating a Kalman filter based approach to overcome such problems.

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] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *Medical Imaging, IEEE Transactions on*, vol. 22, pp. 986-1004, 2003.
[2] B. Pompe, P. Blidh, D. Hoyer, and M. Eiselt, "Using mutual information to measure coupling in the cardiorespiratory system," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 17, pp. 32-39, 1998.
[3] J. W. Fisher, III and J. C. Principe, "Entropy manipulation of arbitrary nonlinear mappings," presented at Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop, 1997.
[4] T. Ikeda, H. Ishiguro, and M. Asada, "Sensor fusion as optimization: maximizing mutual information between sensory signals," presented at Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004.
[5] T. Butz and J.-P. Thiran, "From error probability to information theoretic (multi-modal) signal processing," *Signal Processing Information Theoretic Signal Processing*, vol. 85, pp. 875-902, 2005.
[6] J. W. Fisher, III and J. C. Principe, "A methodology for information theoretic feature extraction," presented at Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on, 1998.
[7] J. W. Fisher and T. Darrell, "Signal level fusion for multimodal perceptual user interface," presented at Proceedings of the 2001 workshop on Perceptive user interfaces, Orlando, Florida, 2001.
[8] J. Fisher, M. J. Wainwright, E. B. Sudderth, and A. S. Willsky, "Statistical and Information-Theoretic Methods for Self-Organization and Fusion of Multimodal, Networked Sensors," *International Journal of High Performance Computing Applications*, vol. 16, pp. 337-353, 2002.
[9] R. Linske, "Self-organization in a perceptual network," *Computer*, vol. 21, pp. 105-117, 1988.
[10] J. W. Fisher, III and T. Darrell, "Speaker association with signal-level audiovisual fusion," *Multimedia, IEEE Transactions on*, vol. 6, pp. 406-413, 2004.
[11] J. W. Fisher, III, "Nonlinear extensions to the minimum average correlation energy filter," in *Electrical and Computer Engineering*: University of Florida, 1997, pp. 183.
[12] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least square tting of two 3-d point sets.," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 9, pp. 698-700, 1987.