Proceedings Paper:
Martinez Hernandez, U, Damianou, A, Camilleri, D et al. (2017) An integrated probabilistic framework for robot perception, learning and memory. In: 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE International Conference on Robotics and Biomimetics (ROBIO 2016), 03-07 Dec 2016 IEEE, pp. 1796-1801. ISBN: 978-1-5090-4364-4.

https://doi.org/10.1109/ROBIO.2016.7866589

An integrated probabilistic framework for robot perception, learning and memory

Uriel Martinez-Hernandez[1], Andreas Damianou[2], Daniel Camilleri[1]
Luke W. Boorman[1], Neil Lawrence[2] and Tony J. Prescott[1]

Abstract— **Learning and perception from multiple sensory modalities are crucial processes for the development of intelligent systems capable of interacting with humans. We present an integrated probabilistic framework for perception, learning and memory in robotics. The core component of our framework is a computational Synthetic Autobiographical Memory model which uses Gaussian Processes as a foundation and mimics the functionalities of human memory. Our memory model, that operates via a principled Bayesian probabilistic framework, is capable of receiving and integrating data flows from multiple sensory modalities, which are combined to improve perception and understanding of the surrounding environment. To validate the model, we implemented our framework in the iCub humanoid robotic, which was able to learn and recognise human faces, arm movements and touch gestures through interaction with people. Results demonstrate the flexibility of our method to successfully integrate multiple sensory inputs, for accurate learning and recognition. Thus, our integrated probabilistic framework offers a promising core technology for robust intelligent systems, which are able to perceive, learn and interact with people and their environments.**

I. Introduction

The aim of social robotics is to better integrate robots into society with the capability to autonomously communicate, learn and assist people. Learning through interaction is critical to social robotics as is the ability to remember past interactions and to use information from what has happened before to make sense of the here and now. The comprehensive integration of data flows, provided by multiple sensory modalities, arising from the environment is an essential component in the design of robust perception and learning algorithms [1]. Despite the advances in robot technology and artificial intelligence methods, these requirements still constitute a great challenge.

Biologically inspired models offer an approach for modelling aspects of the human brain, such as information processing and storage. Cognitive architectures with a learning by observation approach have been implemented to perform assembly tasks and games [2],[3]. These approaches, used in passive mode, do not allow the robotic systems to actively perceive and learn from multiple sources of information. Other approaches have proposed cognitive frameworks that mimic features associated to human memory for attention, learning and natural language processing [4], [5], [6]. The main limitations of these works are their inability to handle multiple sensing modalities, adaptability and lack of memory functions, e.g., compression, chunking and consolidation.

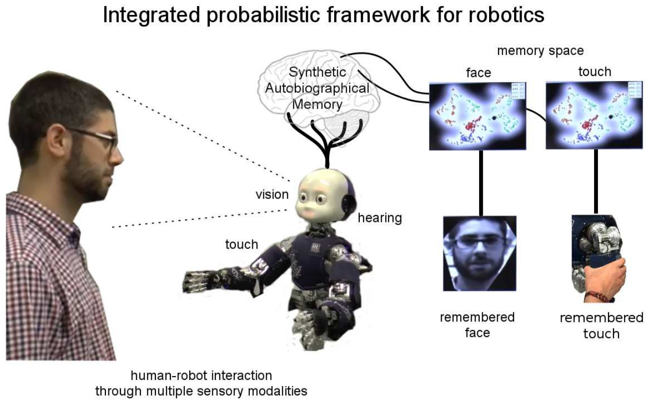


Fig. 1. Integrated probabilistic framework for robot perception, learning and memory. Our framework, based on Synthetic Autobiographical Memory model, allows robots to perceive and learn from multiple sensory inputs.

In this work, we propose an integrated probabilistic framework for robot perception, learning and memory processes. The core of our method is a computational Synthetic Autobiographical Memory (SAM) model, that is analogous to human autobiographical memory and mimics compression, chunking and consolidation functions. The key component of our SAM framework is based on a family of Bayesian latent variable models collectively referred to as a Deep Gaussian process (DGP) [7], [8], [9]. The DGP offers a probabilistic framework for modelling complex data in (un/semi)supervised mode, which has been used as a platform for perception and learning [10], [11].

Perception and learning from multiple sensory modalities, e.g., vision, touch and hearing, are processes also offered by our framework by the use of input *driver* modules. These modules prepare the sensory information in the appropriate format to be understandable by the SAM model. *Driver* modules are also used to handle the output from our framework. This modular approach allows for a robust, adaptable and scalable probabilistic framework, that can be implemented in the context of biomimetic layered control architectures for robot social cognition [12].

*This work was supported by EU Framework project WYSIWYD (FP7-ICT-2013-10) and the European FET Flagship Programme through the Human Brain Project (HBP-SGA1 grant agreement 720270).

[1]Uriel Martinez-Hernandez, Daniel Camilleri, Luke W. Boorman and Tony J. Prescott are with the Sheffield Robotics Laboratory and the Department of Psychology, The University of Sheffield, Sheffield, U.K. uriel.martinez, d.camilleri, l.boorman, t.j.prescott@sheffield.ac.uk

[2]Andreas Damianou and Neil Lawrence are with the Sheffield Robotics laboratory and the Department of Computer Sciences, The University of Sheffield, Sheffield, U.K. andreas.damianou, n.lawrence@sheffield.ac.uk

We implemented our probabilistic framework in the iCub humanoid robot. Our method was validated through perception and learning processes with three human-robot interaction experiments: 1) face recognition, 2) arm movement-based action recognition and 3) touch gesture recognition. For these tasks, vision and touch datasets from human participants, interacting with the iCub, were collected for training and testing our method in off-line and real-time modes respectively. Results show the ability of our method for integration of sensory data, adaptability and accurate perception and learning during human-robot interaction. Furthermore, the principled handling of uncertainty in the perception and prediction stages (through the Bayesian probabilistic formulation) constitutes a promising safety and robust mechanism for real-world robotic applications.

Probabilistic frameworks for perception, learning and memory, inspired by biological systems, offer a coherent path to deploy safe, robust and intelligent systems into society. In this sense, our computational method inspired by human memory, reflects these benefits that are suitable for human-robot interaction and social robotics.

II. Methods

A. *robotic platform*

We used the iCub humanoid robot for the implementation of our integrated probabilistic framework. The iCub is one of the most advanced and open source platforms developed for investigation of cognitive development and human-robot interaction [13]. This robot, that resembles a four year old child, is composed of 53 degrees of freedom and multiple sensory modalities that permit control of robot movements modulated by information from its surrounding environment. The iCub humanoid robot is shown in Figure 1.

Visual data are provided by two RGB cameras with resolution of 640×480 and frame rate of 30fps. These cameras are mounted in the head of the robot, in positions similar to where eyes are located on a human face. This configuration, together with appropriate computer vision algorithms, enables the recognition and tracking of regions of interests from the scene [14]. Physical and safe interaction of the robot with its environment is possible with the use of its tactile sensory system that, built with a capacitive technology, provides pressure measurements in [0 255] sampled at 50 Hz [15], [16]. Previous works on intelligent perception and control have shown that the tactile sensory system of the iCub robot, located in its torso, arms and hands make it capable for exploration, recognition and interaction [17], [18], [19].

Two omnidirectional microphones with noise cancellation feature, from Andrea Electronics, were integrated with the iCub humanoid robot. These microphones, together with Kaldi speech recognition toolkit and text-to-speech modules, allow for verbal communication with humans [20]. This communication method provides a more natural and friendly approach for human-robot interaction.

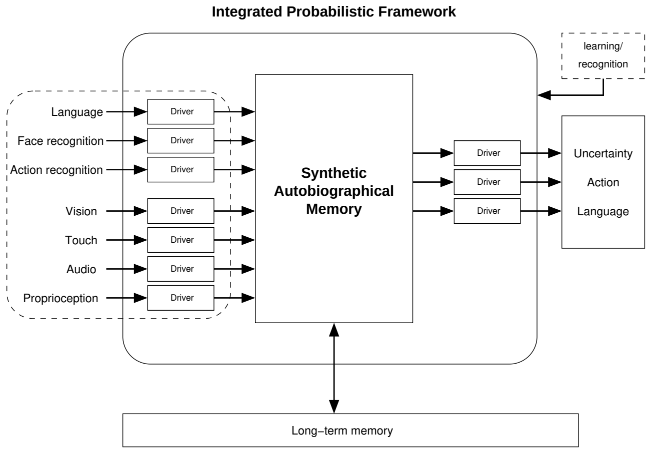


Fig. 2. Integrated probabilistic framework for perception, learning and memory. A Synthetic Autobiographical Memory (SAM) model is the core of our method. Sensory data, from multiple modalities, is preprocessed by driver modules. Knowledge acquired is store in a Long-Term Memory.

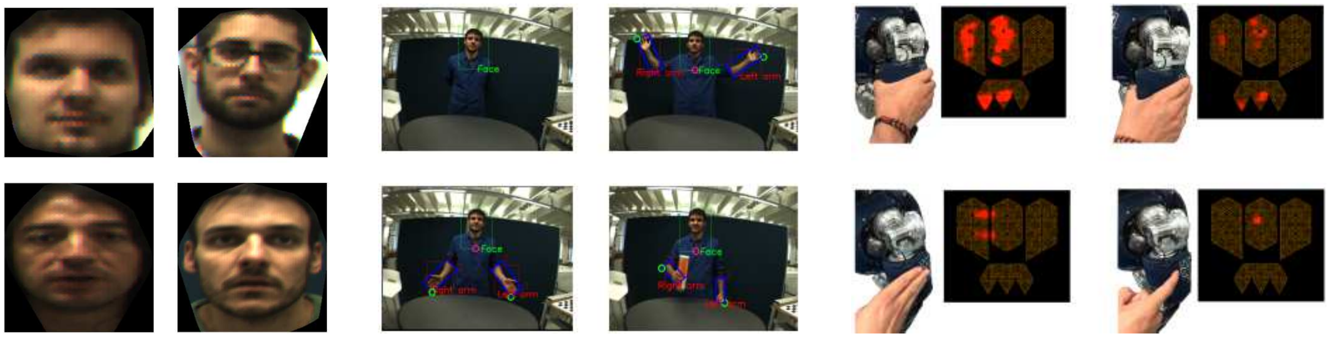
B. *Integrated probabilistic framework*

We propose an integrated probabilistic framework that interfaces with a robotic platform to perform sensation, perception, learning and memory processing. It has been shown that these cognitive processes are highly integrated and essential for behaviour control of biological systems [21]. In our framework, integration is achieved by using the same memory system for learning (storing and recalling) events from the different modalities.

The core component of the framework implements a computational SAM model, that analogously to human autobiographical memory, offers intelligent perception and learning functions. In more detail, the modules that compose our integrated probabilistic framework are shown in Figure 2. The incoming data flow from the multiple modalities and output actions are handled by the *drivers*, which are responsible for data preprocessing. Learned events are consolidated in a long-term memory (LTM) that, developed with a PostgreSQL database, make this information available for subsequent human-robot interaction tasks [22]. The following sections will describe in detail our probabilistic framework.

1) Synthetic autobiographical memory: Our SAM model implements various human memory functionalities such as chunking, compression and consolidation. This approach is inspired by previous studies on hippocampal memory models for navigation in simulated environments [10], [23]. The SAM model is capable to process data flows from multiple sensory modalities, and chunk them into episodes based on the detection of a new event, to improve the understanding of the changing surrounding environment.

Our model takes advantage of the Gaussian process latent variable approach (DGP), which provides a robust probabilistic framework for learning and perception by modelling data at different levels of abstraction [8], [11]. The model represents each sensory stream as a collection of random variables $\mathbf{Y} = \{\mathbf{y}_n\}_{n=1}^{N}$ where $\mathbf{y}_n$ denotes the $n-$th observed instanti-



(a) Face driver (b) Action driver (c) Touch driver

Fig. 3. Driver modules developed for preprocessing of sensory data from multiple modalities, e.g., vision and touch, and sent them to the SAM model. (a) Detection and segmentation of faces. (b) Detection and segmentation of arms for action recognition. (c) Processing of tactile data for recognition of four touch gesture; (top-left) 'hard', (top-right) 'soft', (bottom-left) 'caress', (bottom-right) 'pinch'.

ation. Each data-point $\mathbf{y}_n$ is represented by D features. Real-world perceptual data are very noisy and potentially very high-dimensional, such as raw video signals. To cope with this kind of data, the DGP assumes a simpler, *latent*, low-dimensional (compression) generating space $\mathbf{X} \in \Re^{N \times Q}$ which non-linearly maps to each observation space [24]. The probabilistic construction then defines a density for $p(\mathbf{Y}|\mathbf{X})$ and the learning objective is to learn the posterior for the latent space $p(\mathbf{X}|\mathbf{Y})$ as well as the mapping $f : \mathbf{X} \mapsto \mathbf{Y}$. The generating procedure takes the form:

$$\mathbf{y}_n = f(\mathbf{x}_n) + \boldsymbol{\epsilon}_n, \quad \text{where} \;\; \boldsymbol{\epsilon}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{1}$$

$$\mathbf{x}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{2}$$

where we assume a Gaussian process prior [25] for f, i.e. $f \sim \mathcal{GP}$ and $p(\mathbf{F}|\mathbf{X}) \sim \mathcal{N}$. We denoted by $\mathbf{F} = \{f(\mathbf{x}_n)\}_{n=1}^N$ the instantiations of the function f corresponding to observed data. Notice that we have also assumed normally distributed noise which is corrupting our observations, so that $p(\mathbf{Y}|\mathbf{F}) \sim \mathcal{N}$. By integrating over f one can obtain the desired likelihood density $p(\mathbf{Y}|\mathbf{X})$. The posterior density of the latent space is found through Bayes' rule: $p(\mathbf{X}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})}{\int p(\mathbf{Y}|\mathbf{X})p(\mathbf{X})\mathrm{d}\mathbf{X}}$. The prior latent distribution required in the Bayes' rule can be a fairly uninformative prior (e.g. as in eq. (2)). Alternatively, the DGP framework allows us to use another Gaussian process as a latent prior, to increase expressiveness. This results in a nested definition of the generative model, where equation (2) is now replaced with $\mathbf{x}_n = g(\mathbf{z}_n), g \sim \mathcal{GP}$. Although we described the construction for a two-layer DGP, this nested definition can be recursively applied to obtain a deeper model. The motivation for this construction is that deep architectures are able to capture richer statistical relationships in the data by learning more and more abstraction in each latent layer [26].

The manifestation of functionality analogous to memory is attributed to two elements of the DGP SAM core: firstly, the latent space $\mathbf{X}$ which compresses the observed signal into a non-linear manifold of, typically, reduced dimensionality ($Q \ll D$). Secondly, the inclusion of M anchors $\mathbf{U} \in \Re^{M \times D}$ which are optimized to further compress the signal into a smaller set of variables ($M \ll N$) by being sufficient statistics for the Gaussian process mapping function: $p(\mathbf{F}|\mathbf{X}) = \int p(\mathbf{F}|\mathbf{U}, \mathbf{X})p(\mathbf{U})\mathrm{d}\mathbf{U}$. These anchor points are also referred to as "inducing points" [27] and their role in the memory model is further explained in [10].

Generation of 'fantasy' data, e.g., imagination of novel faces and touch sensations, is also a feature provided by our probabilistic framework. This is implemented by sampling new latent points $\mathbf{X}_*$ and projecting to the observation space through the Gaussian process mapping. This feature offers a tremendous potential for the development of intelligent systems capable of 'imagining' new events, using data stored in memory as humans do. Another advantage stemming from this feature is that it renders the compressed robotic memory fully interactive, rather than being a 'black box'.

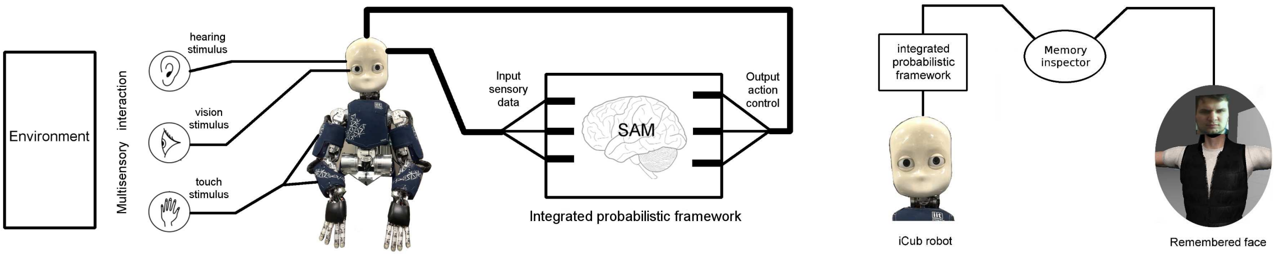
C. Vision sensing

Vision allows robots to visually explore their environment. Here, vision is used for human face and action recognition.

1) Face recognition: Detection of humans in front of the robot is needed to initiate the human-robot interaction. For that reason, the SAM model performs both face recognition and face tracking processes using data from the iCub eyes. Before sending the sensory data to the SAM model, the vision driver shown in Figure 2 performs the following processes using OpenCV: face detection with the Haar Cascade classifier, background subtraction and image resize. This driver is also responsible for the real-time tracking of human faces, which provides a more natural human-robot interaction process and aids face recognition by framing the face within the image area. Figure 3a shows output examples of faces preprocesses and ready for the SAM model.

Preprocessed faces are sent to the SAM model which can be configured for either learning or recognition mode. In learning mode, a model from the set of faces is generated and stored in the LTM. In recognition mode, the test input faces are compared to the already learned faces in the *memory* (i.e. *compressed*) space. The process of receiving vision data for learning and recognition tasks is depicted in Figure 4a.

2) Action recognition: Recognition of human actions, based on arm movements, allows the robot to reduce ambiguity and have a better understanding during a human-robot interaction. The action driver, shown in Figure 2, is



(a) (b)

Fig. 4. (a) Integrated probabilistic framework implemented in the iCub humanoid robot for perception, learning and memory processes. The robot receives stimuli from multiple sensory modalities through the interaction with humans and its surrounding environment. Sensory data are analysed by the SAM model and resultant actions are sent back for robot control. (b) Memory inspector module that allows to observe the current state of the SAM model. Thus, it is possible to know what human face is being remembered or 'imagined' by the robot.

responsible for detection of human arms and their movements captured by the iCub eyes. This driver uses a colour-based filter for detection and segmentation of human arms. Changes in $x-, y-$, and $z-$positions of human arms are used to detect their movements. Here, we defined five human actions: waving, lift up, put down, push and pull. Figure 3b shows examples of human actions with our SAM model.

Then, the action driver sends the sequence of processed actions (arm movements) to our probabilistic framework, which allows the iCub to learn and recognise different actions through the observation of human participants.

D. Touch sensing

Touch is an important sensing modality that, together with intelligent perception and control methods, allows robots to safely and autonomously explore and build a physical representation of their surrounding environment [28], [29]. For that reason, we developed a touch driver for handling tactile data and send them to our probabilistic framework for learning and recognition of touch gestures (Figure 2). This driver captures the raw pressure measurements from the tactile sensors of the iCub based on a human-robot interaction. Tactile interactions are detected and segmented into contact events, using a thresholding approach configured with the minimum pressure detected by the skin of the iCub. Here, we used four types of tactile gestures, labelled as "hard", "soft", "caress" and "pinch", based on the pressure value and duration of touch employed by humans on the iCub humanoid robot. Examples of these tactile gestures are shown in the Graphical User Interface (GUI) in Figure 3c.

The output from the touch driver is sent to our SAM model for learning and recognition of touch gestures. In learning mode, touch gesture models are generated by our framework and stored in the LTM. In recognition mode, the data from touch gestures are compared with previously learned models, obtaining the most probable label for the applied touch.

E. Memory inspector

We have implemented a *memory inspector* module in our integrated probabilistic framework. This module is aimed at visually displaying the content of the robot memory, to observe what the robot is 'imagining' or 'remembering' at a specific time during a human-robot interaction. This functionality highlights the advantage of the generative Bayesian formulation upon which the SAM core is built. Currently, the memory inspector is capable of describing a memory and creating a decomposed representation of an action and agent, displaying them within a simulation environment, which mimics a memory playback. Figure 4b shows an example of a 'fantasy' face, remembered by the SAM model and displayed by the memory inspector, within the virtual environment embodied by a virtual agent. Generation of human arm movements and touch gestures are also features to be included in the virtual agent in future work.

III. Results

Validation of our integrated probabilistic framework was based on experiments for learning and recognition of faces, arm movements and touch gestures. These experiments were performed in real-time through human-robot interaction.

To train our proposed framework we collected multiple datasets from vision and touch sensing. Visual data were collected for the detection of faces and arms from three human participants located in front of the iCub humanoid robot. Faces and arm movements at different orientations and positions inside the visual scene of the robot were captured to improve the learning process. Tactile data were collected by applying multiple touch gestures to the arms of the iCub humanoid robot by human participants. The sensory data were processed by their corresponding drivers and sent to the SAM model, initially configured in learning mode, to build models for faces, actions and touch, and store them in the LTM built with a PostgreSQL database. Figure 5 shows an example of the faces memory model built from the training data. Real-time experiments, described in the following sections, were performed to test our probabilistic framework both quantitatively (accuracy in recognition) and qualitatively (functional properties of the framework, seamless embodiment and interaction with the iCub robot).

A. Multisensory human-robot interaction

We tested the accuracy and multisensory capability of our integrated probabilistic framework for learning and recognition of faces, action and touch gestures. This process was performed using multisensory data, obtained in real-time mode, from humans interacting with the iCub robot.

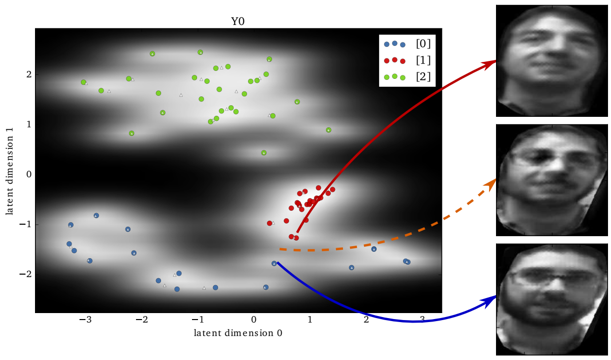


Fig. 5. Example of memory space for learning and recognition of human faces. Each point in the 2D point space represents the compressed representation of the high-dimensional facial image (shown on the right) perceived during training. The memory space is forming three clusters (coloured distinctively), which signifies successful learning, since three different individuals were presented to the robot. The background gray shading quantifies the uncertainty of recalling a face from each respective point in the continuous 2D memory space. Recalling a face from in-between the training points results in the generation of a 'fantasy' face, as demonstrated with the dotted line.

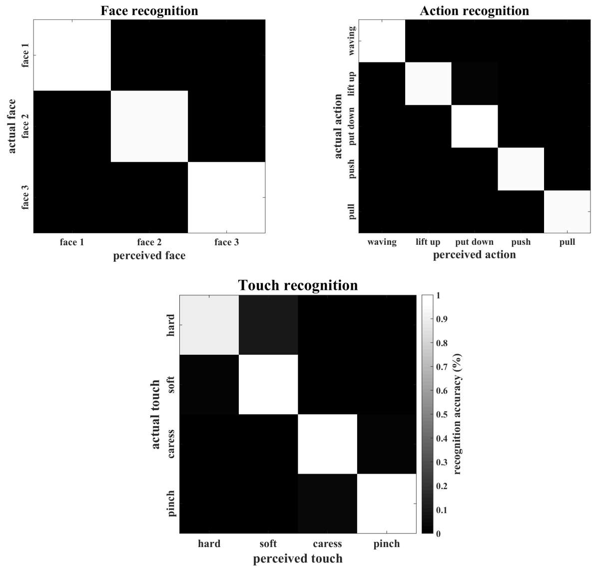


Fig. 6. Confusion matrices for face, action and touch recognition. (top-left) Human face recognition from three participants, achieving an accuracy of 99.0%. (top-right) Five arm-based human actions were performed with recognition accuracy of 98.0%. (bottom) Four types of touch gesture were recognised with an accuracy of 95.0%.

1) Face perception: Real-time recognition of human faces was performed to validate our integrated probabilistic framework with data from vision sensing. For this experiment, we asked human participants to stand in front of the robot, one at the time. The robot was able to detect, segment and track human faces with the vision driver. Then, the processed vision data were sent to the SAM model for face recognition in real-time mode. The recognition accuracy of our method was tested with faces from various human participants (see Figure 3a), that were tracked and captured at various orientations, not only looking straight to the robot.

The confusion matrix in Figure 6a shows the results for face recognition of three human participants. These results were obtained from 20 repetitions of the experiment for each participant, while 150 face samples were used for training our method. A total accuracy of 99.33% was obtained by our framework, where individual accuracies of 100%, 98% and 100% were achieved by each participant.

2) Action perception: For testing of human actions, based on arm movements, we asked human participants to stand in front of the robot and do the following set of actions: 'waving', 'push', 'pull', 'lift up' and 'put down'. These actions were performed multiple times in a random order. Similar to recognition of faces, data from human actions were captured with different backgrounds and light conditions to test the robustness of our probabilistic framework.

The action driver allowed the iCub humanoid robot to detect, segment and track arms from human participants in real-time mode. Processed data were sent to the SAM model, configured in recognition mode. This experiment was repeated 20 times by each participant. We used 150 samples for training our method. Results from these experiments are shown in the confusion matrix of Figure 6b, where a total accuracy of 98.40% was obtained by our integrated probabilistic framework. Recognition accuracies of 100%, 97%, 100%, 98% and 97% were achieved for 'waving', 'lift up', 'put down', 'push' and 'pull' arm movements.

3) Touch perception: We tested the capability of the iCub robot for recognition of different touch gestures during a human-robot interaction process. For this purpose, multiple touch gestures labelled as 'hard', 'soft', 'caress' and 'pinch', were applied by human participants on the skin located in the arms of the iCub humanoid robot.

For each tactile interaction, the touch driver detected the pressure applied, by human participants, on the skin of the robot, segmented the contact detected, and sent the prepared data to the SAM model, configured in recognition mode. The robot was able to accurately recognise the different touch gestures and provide a label (learned in training phase). The test process was repeated 20 times by each participant, while 150 touch samples were used for training our method. Recognition accuracy for touch gesture is shown by the confusion matrix in Figure 6c. High accuracies of 97.0%, 97.0% and 96.0% were achieved by 'soft', 'caress' and 'pinch' gestures respectively, while 'hard' gesture achieved an accuracy of 90.0%. Overall, an accuracy of 95% was achieved, allowing the robot to accurately perceive different touch gestures during a human-robot interaction process.

Recognition results from face, action and touch, not only demonstrate the accuracy and robustness of our integrated probabilistic framework, but also its ability to process multisensory data from human-robot interaction in real-time.

IV. Conclusion

We have presented an integrated probabilistic framework for robot perception, learning and memory processes through the interaction with humans. Our multicomponent framework offers a robust and scalable approach that integrates multiple modality information for better learning in robotics.

A biologically inspired computational Synthetic Autobiographical Memory (SAM) model, that implements compression, chunking and consolidation human memory functionalities, was developed as the core of our framework. The SAM model uses a Bayesian formulation based on a Gaussian process latent variable framework, which offers intelligent perception and learning at different level of abstractions. Integration and processing of vision, touch and hearing sensing modalities, which are highly unified and essential for robust perception and learning, are available in our implemented framework. Scalability of our approach is based on the implementation of driver modules, which preprocess and prepare the sensory data in appropriate formats, making the sensory input a transparent process for the SAM model. Scalability with regards to handling larger amounts of data is a promising research direction which we wish to pursue in the future, and is expected to open up new ways of human-robot interaction, e.g. learning with millions of video frames.

We used the iCub robot for validation of our integrated probabilistic framework. For training, we collected multiple datasets from vision and touch using the eyes and skin of the robot. For testing our method in real-time mode, human participants were asked to interact with the robot, performing different human arm movements and applying different touch gestures on the skin of the robot. Results showed the ability of the robot to recognise and 'imagine' human faces, arm movements and touch gestures with high accuracy.

Deployment of intelligent robots capable of perceive, learn and safely interacting with humans remains a challenge for scientists and engineers. However, relevant progress in robot perception and learning achieved by biologically inspired probabilistic models, like our method proposed in this work, give a promising route for development and integration of safe and intelligent robots in society.

Acknowledgements: We thank the developers of GPy for open-sourcing the software and Max Zwiessele for assisting with the code.

REFERENCES

[1] D. J. Lewkowicz and A. A. Ghazanfar, "The emergence of multisensory systems through perceptual narrowing," *Trends in cognitive sciences*, vol. 13, no. 11, pp. 470–478, 2009.

[2] A. L. Thomaz and M. Cakmak, "Learning about objects with human teachers," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 2009, pp. 15–22.

[3] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Learning by watching: Extracting reusable task knowledge from visual observation of human performance," *Robotics and Automation, IEEE Transactions on*, vol. 10, no. 6, pp. 799–822, 1994.

[4] E. M. Altmann and J. G. Trafton, "Memory for goals: An activation-based model," *Cognitive science*, vol. 26, no. 1, pp. 39–83, 2002.

[5] J. R. Anderson, L. M. Reder, and C. Lebiere, "Working memory: Activation limitations on retrieval," *Cognitive psychology*, vol. 30, no. 3, pp. 221–256, 1996.

[6] D. Hart and B. Goertzel, "Opencog: A software framework for integrative artificial general intelligence," *Frontiers in Artificial Intelligence and Applications*, vol. 171, p. 468, 2008.

[7] N. D. Lawrence and A. J. Moore, "Hierarchical Gaussian process latent variable models," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 481–488.

[8] A. Damianou and N. Lawrence, "Deep Gaussian processes," in *Proceedings of the Sixteenth International Workshop on Artificial Intelligence and Statistics (AISTATS)*, ser. AISTATS '13, C. Carvalho and P. Ravikumar, Eds. JMLR W&CP 31, 2013, pp. 207–215.

[9] A. Damianou, "Deep Gaussian processes and variational propagation of uncertainty," *PhD Thesis, University of Sheffield*, 2015.

[10] A. Damianou, C. H. Ek, L. Boorman, N. D. Lawrence, and T. J. Prescott, "A top-down approach for a synthetic autobiographical memory system," in *Biomimetic and Biohybrid Systems*. Springer, 2015, pp. 280–292.

[11] A. Damianou, C. Ek, M. Titsias, and N. Lawrence, "Manifold relevance determination," in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*. New York, NY, USA: Omnipress, July 2012, pp. 145–152.

[12] S. Lallee and P. F. Verschure, "How? why? what? where? when? who? grounding ontology in the actions of a situated social agent," *Robotics*, vol. 4, no. 2, pp. 169–193, 2015.

[13] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. Von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, *et al.*, "The icub humanoid robot: An open-systems platform for research in cognitive development," *Neural Networks*, vol. 23, no. 8, pp. 1125–1134, 2010.

[14] M. Petit, S. Lallée, J.-D. Boucher, G. Pointeau, P. Cheminade, D. Ognibene, E. Chinellato, U. Pattacini, I. Gori, U. Martinez-Hernandez, *et al.*, "The coordinating role of language in real-time multimodal learning of cooperative tasks," *IEEE Transactions on Autonomous Mental Development*, vol. 5, no. 1, pp. 3–17, 2013.

[15] U. Martinez-Hernandez, "Tactile sensors," in *Scholarpedia of Touch*. Springer, 2016, pp. 783–796.

[16] A. Schmitz, M. Maggiali, L. Natale, B. Bonino, and G. Metta, "A tactile sensor for the fingertips of the humanoid robot icub," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 2212–2217.

[17] U. Martinez-Hernandez, T. J. Dodd, M. Evans, T. J. Prescott, and N. F. Lepora, "Active sensorimotor control for tactile exploration," *Robotics and Autonomous Systems*, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.robot.2016.09.014

[18] U. Martinez-Hernandez, N. F. Lepora, and T. J. Prescott, "Active haptic shape recognition by intrinsic motivation with a robot hand," in *World Haptics Conference (WHC), 2015 IEEE*. IEEE, 2015, pp. 299–304.

[19] U. Martinez-Hernandez, A. Rubio-Solis, and T. Prescott, "Bayesian perception of touch for control of robot emotion," in *International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016.

[20] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The kaldi speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, Dec. 2011, iEEE Catalog No.: CFP11SRW-USB.

[21] M. S. Cohen, S. M. Kosslyn, H. C. Breiter, G. J. DiGirolamo, W. L. Thompson, A. Anderson, S. Bookheimer, B. R. Rosen, and J. Belliveau, "Changes in cortical activity during mental rotation a mapping study using functional mri," *Brain*, vol. 119, no. 1, pp. 89–100, 1996.

[22] G. Pointeau, M. Petit, and P. F. Dominey, "Successive developmental levels of autobiographical memory for learning through social interaction," *IEEE Transactions on Autonomous Mental Development*, vol. 6, no. 3, pp. 200–212, 2014.

[23] L. W. Boorman, A. Damianou, U. Martinez-Hernandez, and P. T.J., "Extending a hippocampal model for navigation around a maze generated from real-world data." in *Lecture Notes in Artificial Intelligence (Living Machines), Biomimetic and Biohybrid Systems*, 2015.

[24] M. K. Titsias and N. D. Lawrence, "Bayesian Gaussian process latent variable model," in *International Conference on Artificial Intelligence and Statistics*, 2010, pp. 844–851.

[25] C. E. Rasmussen, "Gaussian processes for machine learning," 2006.

[26] Y. Bengio, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[27] J. Quinonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *The Journal of Machine Learning Research*, vol. 6, pp. 1939–1959, 2005.

[28] U. Martinez-Hernandez, T. Dodd, T. J. Prescott, and N. F. Lepora, "Active bayesian perception for angle and position discrimination with a biomimetic fingertip," in *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*. IEEE, 2013, pp. 5968–5973.

[29] U. Martinez-Hernandez, G. Metta, T. J. Dodd, T. J. Prescott, L. Natale, and N. F. Lepora, "Active contour following to explore object shape with robot touch," in *World Haptics Conference (WHC), 2013*. IEEE, 2013, pp. 341–346.