

ПРОБЛЕМА ВИЗУАЛЬНОЙ КОММУНИКАЦИИ В СОЦИАЛЬНОЙ РОБОТОТЕХНИКЕ¹

В.А. Сербин

Национальный исследовательский Томский государственный университет

Рассматриваются современные подходы к проектированию визуальных сенсоров социальных роботов. Проводится сравнение концепций зрения в робототехнике и гуманитарных науках. Выделены актуальные направления работы проектирования оптических сенсоров и моделей зрения для социальных роботов. Актуализируются основные проблемы, существующие на данном этапе исследований и разработки систем искусственного зрения.

Ключевые слова: социальная робототехника, взаимодействие человека и робота, визуальная коммуникация.

VISUAL COMMUNICATION PROBLEM IN SOCIAL ROBOTICS

V.A. Serbin

National Research Tomsk State University

Article represents modern approaches in construction of visual sensors of social robots. Comparison of engineering vision concepts with humanitarian vision concepts are being provided. Highlighted current trends in the design of optical sensors and models for social vision. robots. Updated the main problems existing at the present stage of research and development of artificial vision systems.

Key words: social robotics, human-robot interaction, visual communication.

Одним из актуальных вопросов социальной робототехники является изучение критериев социальности роботов, т.е. тех черт робота, которые позволили бы назвать его социальным агентом. Так как «социальность» отчасти проявляется в наших поведенческих моделях, то социальный робот должен адекватно реагировать на изменения окружающей обстановки и действия человека. В связи с этим встает вопрос о технической реализации восприятия и интерпретации роботом визуальных, звуковых, тактильных и прочих раздражителей. В статье мы рассмотрим некоторые концептуальные пробле-

¹ Данная статья написана при поддержке гранта РФФИ 12-06-33047 – «Исследования междисциплинарных научных оснований социальной робототехники в контексте гуманитарной информатики».

мы создания систем распознавания визуальных образов, спроектированных для социальных роботов.

Под визуальной коммуникацией мы будем понимать возможность контекстуальной интерпретации роботом информации, воспринимаемой посредством визуальных сенсоров. Робот должен адекватно воспринимать как пространство, в котором разворачивается событие, так и действия людей, основываясь на информации, предоставляемой аналогом зрительного аппарата. Специфика требований, предъявляемых к тому, как и что видит социальный робот, заключается в том, что восприятие роботом мира должно быть приближено к восприятию мира человеком, иначе говоря, робот должен видеть примерно так же, как и мы.

Следует понимать, что решение роботом интеллектуальных задач, которые мы воспринимаем как простые и очевидные, требует комплексных инженерных и программных решений, а воспроизводимые роботом реакции, определяемые как социальные, нуждаются в проработке со стороны специалистов в области гуманитарных наук. В рамках социальной робототехники как научного направления происходит интеграция разнородных наук и дисциплин, что открывает множество путей для решений одних задач и проблем. Многообразие этих решений сказывается на том, как разработчики конкретных решений (моделей роботов) подходят к проблеме восприятия роботом визуальной информации.

На данный момент главными достижениями в развитии сенсоров роботов можно назвать визуальные – отвечающие за «зрение» робота, способность распознавать человека, группы людей, жесты и др., и звуковые – отвечающие за такие способности робота, как слух, интерпретация речи, обнаружение источника звука. Отдельно следует выделить разработки тактильных сенсоров, которые позволяют роботам ощущать давление, оказываемое на поверхность корпуса робота, а также регулировать силу сжатия манипуляторов (рук, клешней, захватов и подобных им инструментов) [1]. Существуют также «экзотические» сенсоры, имитирующие инструментарий биологических форм жизни, например усики [2].

При том, что зрительная сенсорика робота, вне зависимости от его формальных характеристик (гуманоид, антропоморф, зооморф), копирует функции зрительного аппарата человека или животного, условные «глаза» имитируются комбинацией нескольких технологий, аппаратных и программных.

Важнейшим и специфическим для социальной робототехники аспектом разработки визуальных сенсоров является то, что каждый используемый сенсор напрямую влияет на продуктивность взаимодействия человека и робота, т.е. сенсор одновременно является интерфейсом. Если мы рассматриваем сенсорику робота с позиций социальной робототехники, то необходимо учесть двойственный характер сенсора: его техническую часть – то, что устройство позволяет «почувствовать» роботу, и социально-коммуникативную – то, как человек воспринимает действия робота как социального агента.

Каждая конкретная реализация концепции зрения социального робота направлена на решение очевидной задачи – обеспечение робота набором программно-аппаратных средств, предоставляющих возможность получать информацию о происходящем вокруг. Практика нам показывает, насколько эта задача сложна и нетривиальна. Далее на примерах разработок крупных научных организаций мы опишем ряд проблем, возникающих в области создания визуальных сенсоров.

Инженеры Personal Robot Group (Массачусеттский технологический институт) предложили концепцию Активного зрения для робота в начале 2000-х годов. Концепция изначально была воплощена в прототипе Kismet [2], а позже – в нескольких моделях социальных роботов-компаньонов (Huggable™ [1], Leonardo [4], MDS [5]). Авторы исходят из положения о том, что для осмысленного взаимодействия человека и робота необходимо, чтобы они обладали схожими инструментами восприятия действительности. Зрение является одним из таких инструментов. При этом каждый из участников такого взаимодействия должен догадываться о том, что другой думает и чувствует (видит). Зрение является одним из важнейших для социальных коммуникаций органом чувств, поэтому, считают разработчики системы активного зрения, в социальной робототехнике необходимо имитировать человеческие органы зрения [6. Р. 443].

Под активным зрением понимается комплекс технологий, решающих несколько задач:

- 1) визуальное распознавание объектов роботом;
- 2) идентификация объектов, человека или людей;
- 3) реакция на визуальные раздражители;

4) социальное поведение, заключающееся в коммуникации посредством взгляда, примитивной мимики и соблюдения «личного пространства».

Особенностью решения, предложенного в МИТ, является рассмотрение инженерных вопросов робототехники в совокупности с социально-коммуникативными. Система активного зрения внешне представляет собой имитацию глаз с подвижными веками и бровями. Камеры, обеспечивающие «зрение» роботу, по аналогии с устройством человеческого глаза находятся внутри искусственного глаза. Одной пары подвижных камер оказалось мало для обеспечения робота достаточной визуальной информацией для формирования его социального поведения, поэтому разработчики установили две статичные камеры с более широким углом обзора. Две пары камер выполняют различные функции, статичные фиксируют общую обстановку, подвижные направляются на объект интереса робота.

Об улучшении социальных и коммуникативных функций робота с системой активного зрения позволяет судить эксперимент, заключавшийся в «знакомстве» нескольких человек с прототипом робота. Перечислим достижения авторов разработки [Там же. Р. 452].

- Робот способен задавать персональное пространство между собой и человеком: если человек подходил слишком близко, робот изображал раздражение и пятился назад. Важно то, что человек, нарушающий дистанцию общения с роботом, правильно интерпретировал его реакцию и отодвигался на некоторое расстояние.

- Робот способен привлекать людей к взаимодействию мимикой и действиями.

Стоит отметить то, что действия, продуцируемые роботом при помощи системы активного зрения, носят утилитарный характер. Отдаление от человека, слишком сильно приблизившегося к голове робота, необходимо, так как камеры не могут сфокусироваться на объекте, а привлечение внимания человека, находящегося слишком далеко, необходимо для идентификации объекта (человека в данном примере). Социальное поведение компенсирует аппаратные изъяны используемых камер, способных к передаче картинки среднего качества.

Главными недостатками системы активного зрения на момент ее создания авторы называли невозможность передачи роботу информации об объеме и дистанции по визуальному каналу, невозможность ориентации по звуку (робот мог фокусироваться на одном человеке, который к нему обращался), и в качестве перспективы развития робозрения было предложено использовать больше камер. Камеры позволяют роботу создать карту пространства с людьми или объектами, с которыми возможно взаимодействие.

Используя цифровую камеру в качестве оптического сенсора, мы получим двухмерное изображение, из которого робот не сможет извлечь информацию о перспективе и положении объектов в пространстве. Ориентация робота в пространстве – сложнейший вопрос робототехники; сравнительно недавно появилось решение этой проблемы, позволяющее использовать оптические сенсоры, – технология RGB-D.

RGB-D позволяет совмещать традиционную цветовую информацию (RGB) с информацией о глубине и расстоянии (D – depth (англ), глубина). Подобные сенсоры используются в игровом контроллере для XBOX и PC – Microsoft Kinect. Есть модификация Kinect для исследовательской работы, часто используемая в робототехнике.

RGB-D – открытая исследовательская платформа, использование которой в чистом виде не даст результата. В робототехнике RGB-D используется как дополняющая традиционные оптические сенсоры технология, иногда в сочетании с 3d-моделями объектов.

Мы назовем несколько апробированных сценариев использования RGB-D сенсоров в социальной робототехнике.

В Корнельском университете (США) предлагают использование RGB-D-сенсора (в сочетании с технологией, позволяющей обнаруживать движения человеческого скелета, – Openni's skeleton tracker [7]) [8]. Целью авторы называют описательное маркирование комплексных и продолжительных действий человека, состоящих из ряда более мелких действий. Под маркированием понимается создание виртуальной модели действия, которая бы позволила роботу обнаруживать, идентифицировать и запоминать действия человека, наблюдая за ним [Там же. Р. 1]. В рамках эксперимента рассматривалось несколько бытовых сценариев действий человека, таких как приготовление поп-корна, перемещение кувшина, наливание молока в чашку. Авторы предложили использование этой технологии в сервисной робототехнике, таких сценариях [Там же. Р. 14]:

- после того как человека принял пищу, робот убирает стол;
- робот обнаруживает, что человек принимает лекарство, и приносит стакан воды;
- робот обнаруживает, что человек готовит завтрак, и помогает, убирая продукты в холодильник.

Нужно отметить, что робот обнаруживает и распознает ситуацию, используя «зрение», точнее – комбинацию оптического сенсора и RGB-D-сенсора. В перспективе эта технология должна помочь в

создании самообучающихся роботов, идентифицирующих действия человека посредством наблюдения.

В Университете Фрайбурга, Германия, RGB-D-сенсор предлагают использовать для улучшения навыков робота в распознавании объектов [9]. Исследование немецких ученых направлено на поиск оптимального алгоритма сочетания традиционной RGB информации – двумерного изображения с данными о глубине и объеме. Фактически любая модель визуального сенсора, состоящая из нескольких сенсоров и использующая несколько алгоритмов анализа их показаний, является мультимодальным детектором. Авторы настаивают, что именно такие детекторы обеспечивают наилучший результат в визуальном обнаружении объектов, в том числе людей [Там же].

Назовем задачи социальной робототехники, в решении которых могут помочь RGB-D-сенсоры:

- Создание самообучающихся роботов, копирующих действия человека.
- Разработка сенсорных систем, идентифицирующих человека и группы людей.
- Создание моделей действий человека с целью подражания или поддержки со стороны робота.

Отдельное место в развитии способностей зрения социальных роботов занимает проблема взгляда. Вслед за разработчиками визуальных сенсоров для роботов мы отмечали специфику зрения как важной части коммуникативных связей между людьми: взгляд, его направление, прищур глаз и т.д.

Исследователи из Лаборатории социальной робототехники Йельского университета (США) рассматривают вопросы невербальной коммуникации человека и робота, основанной на взгляде и движениях глаз [10]. Главной задачей в этом направлении является разработка такого робота, взгляд которого человек бы воспринимал как осмысленный и мог бы правильно интерпретировать «жесты», производимые глазами робота.

В конце 1990-х преобладающей моделью зрительных систем роботов были обобщающие вычислительные системы, которые, по словам сотрудников лаборатории, являлись грубыми попытками имитации человеческого зрения и способов коммуникации посредством глаз [Там же. Р. 159]. В таких моделях цельная сцена редуцируется к минимальному количеству точек, определенные сочетания

которых провоцируют обратную реакцию робота, т.е. каждому сочетанию дается описание и привязывается какое-либо действие. Сцены были статичными, т.е. происходил покадровый анализ поступающего потока видео. Система зрения, выделяющая в кадрах определенные точки интереса, крайне ограничена и не соответствует множеству моделей видения, представленных в биологической и культурной реальностях.

Мы можем сравнить первые попытки создания моделей интерпретации визуальных сообщений с тем, как в рамках структурализма и постструктурализма разрабатывалась методология анализа кинематографа [11–13]. Сравнение уместно до тех пор, пока мы приравниваем визуальный ряд, передаваемый роботу визуальному ряду кинофильма: набор движущихся кадров разбирается на элементы (знаки), различные сочетания которых производят некоторое значение. Показательным будет то, как Ролан Барт строил анализ фильмов на фотограммах – распечатанных кинокадрах [12, 13]. Подобно Барту и другим представителям семиотики кино, инженеры-робототехники предположили, что можно каждый конкретный кадр редуцировать к минимальному количеству ключевых знаков, появление которых в поле зрения робота должно спровоцировать какие-либо действия или вычисления с его стороны. Мы можем сказать, что семиотики кино и разработчики алгоритмов интерпретации видеоряда «глаз» робота столкнулись со схожей проблемой в различные промежутки времени. Эта проблема заключалась в неспособности описать значения, генерируемые с течением времени: понимание зримого образа человеком сопряжено с тем, как этот образ изменяется. В семиотике кино проблема была выражена различными терминологическими конструкциями: «третий смысл» [13], «третье членение кода» [11] и др. Техническое решение проблем предшествующей методологии систем распознавания визуальных образов заключалось в создании двухмерной системы координат, состоящей из времени и пространства.

Важно понимать, что для имитирования социального поведения при использовании при помощи такой модели необходимо описать огромное количество поведенческих клише, которые бы воспроизводил робот. Если мы учтем многообразие и ситуативность коммуникативных практик человека, то создание подобной базы данных представляет собой практически невыполнимую задачу.

В качестве альтернативы предлагается создание такой системы взгляда и фокусирования внимания социального робота, которая бы как можно точнее имитировала движения глаз человека [10. Р. 160]. Чтобы успешно эмулировать коммуникацию, опосредованную взглядом, необходимо знать то, как она осуществляется между людьми. Авторы приходят к положению о том, что статичное распознавание каких-либо объектов внутри сцены (кадра) никак не соответствует действительности (т.е. тому, как человек интерпретирует мимику, взгляды и жесты другого), так как в процессе формирования значения одну из главных ролей играет время, а точнее – то, как именно изменится положение объекта в пространстве с течением времени. Для преодоления проблем предшествующей методологии систем распознавания визуальных образов, не учитывающей семиозиса, происходящего во временной длительности движения, было предложено создать двухмерную систему координат, состоящую из времени и пространства.

Исследование, проведенное в Йельском университете, интересно также сравнением моделей зрения человека с искусственными моделями. Данные, полученные с eye-tracker (устройством, отслеживающим направление взгляда), сравнивались с моделями вычисления зрительного внимания (computation of visual attention) роботами. Исследовательская группа планировала скорректировать математические алгоритмы, используя данные наблюдений за глазом человека. В результате выяснилось, что зрительное «поведение» человека носит преимущественно ситуативный характер. Таким образом, создание какой-то общей модели зрения и взгляда для социальных роботов пока затруднительно, но в рамках определенных сценариев это вполне возможно [Там же. Р. 171].

В Йельском университете было проведено исследование, не относящееся напрямую к проектированию сенсоров для социальных роботов, но затрагивающих проблемы самоидентификации робота и его ориентации в пространстве при помощи зрения. Оригинальность подхода заключается в использовании зеркала. Тестовый робот обращался к зеркальному отражению для изучения пространственных отношений между собственными манипуляторами и окружающими его объектами [15]. Общей целью исследовательского проекта было создание робота, способного пройти зеркальный тест, разработанный психологом Гордоном Галлапом (Gordon Gallup) в 1970 г. Тест проверяет способность животного к самоидентификации себя с от-

ражением в зеркале. Прикладной целью проекта было улучшение системы позиционирования манипуляторов робота в пространстве. Подобно ребенку, составляющему образ своего тела из отражения в зеркале, робот должен был сравнить требуемое положение «руки» с тем, каким оно получилось де-факто, в чем ему должно было помочь зеркальное отражение.

Исследование не дало конкретных результатов: анализ отражения в зеркале оказался довольно затруднителен, и робот показывал лучшие результаты в постановке манипуляторов, опираясь на собственное зрение. Но, как считают авторы проекта, у технологии, использующей зеркальное отражение, есть нераскрытый потенциал [15. Р. 1995–1996].

Сама идея о построении идентичности посредством зеркала происходит из психологии и психоанализа. Пример интересен инженерным переносом гуманитарных концепций формирования личности человека на робота. Инженерная адаптация оказалась достаточно сложным процессом, на котором сказалась невозможность понимания роботом пространства. Жак Лакан отводил одну из ключевых ролей в формировании образа целостного «Я» человека зеркальному отражению [14. С. 509–516]. Согласно Лакану, без внешнего по отношению к человеку медиума социальное поведение невозможно, так как мы всегда располагаем увиденным образом себя и опираемся на него, выстраивая коммуникацию с другими людьми. Мы можем предположить, что воспроизведение на программном уровне некоторых механизмов психики, отвечающих за восприятие собственного образа, заложит основу для усовершенствования существующих алгоритмов интерпретации визуального ряда роботом.

Стоит заострить внимание на том, что многие разработчики социальных роботов не ставят целью копирование социальных клише и стереотипов, а пытаются создать модель или образ поведения, которое люди воспримут как социальное. В этом контексте уместно вспомнить машину Шеннона, которая имитировала английскую речь, зная частотность употребления букв, диграфов и тетраграфов [16]. Люди, слышавшие синтезированную речь (буквально – поток звуков), узнавали и различали слова, т.е. воспринимали ее как осознанную. Эксперимент подводит нас к выводу о том, что человек склонен интерпретировать некоторые явления как социальные (в пример можно привести то, как люди характеризуют поведение домашних животных, приписывая им человеческие качества), при ус-

ловии, что они напоминают ему таковые. Вероятно, одной из задач в разработке визуальных сенсоров социальных роботов в ближайшее время будет поиск свойственных человеку способов невербальной коммуникации и последующая их формализация в виде программно-аппаратных решений.

Подведем итоги. На сегодня в области проектирования оптических сенсоров и моделей зрения для социальных роботов актуальны следующие направления работы:

- Исследования в области технологий распознавания объектов. Под объектами понимаются люди, группы людей, предметы и т.д.
- Формирование аналога человеческого восприятия объема и перспективы у роботов. Главной задачей здесь является имитация функций зрительного аппарата человека, в робототехнике эта функция частично выполняется комплексными сенсорами.
- Визуальная идентификация поведения и действий человека.
- Разработка моделей взгляда и невербальной коммуникации для социальных роботов.
- Ориентация робота в пространстве при помощи зрения.
- Создание программных алгоритмов распознавания визуального образа.

Отметим, что ставя задачу имитации человеческого зрения, разработчикам приходится конструировать сенсоры и сочетать их так, чтобы информация, ими поставляемая, была избыточна. В пример можно привести сочетание нескольких камер и RGB-D-сенсора. Роботу для успешной имитации социального поведения необходимо «видеть» иначе и больше, чем человек.

Среди проблем, существующих на данном этапе исследований и разработки систем искусственного зрения, можно выделить следующие:

- Невозможность создания общей системы зрения социального робота, которая позволила бы ему успешно адаптироваться к любой ситуации. Сегодня возможно ее частичное решение через создание наиболее частых сценариев, в которых социальный робот будет задействован.
- Трудности в идентификации окружения, в котором находится робот. Это часть комплексного вопроса о способах навигации определения себя в пространстве социального робота.

В завершение мы отметим то, что во многих реализациях систем зрения не хватает фундаментальной концепции, которую можно бы-

ло бы позаимствовать из областей психологии, социальной философии, культурологии и др. Для социальной робототехники принципиально важным остаются вопросы не только о том, как видит человек и как зрительный аппарат человека можно симитировать, но и то, как мы коммуницируем друг с другом посредством жестов, мимики глаз и движений тела. Все это, помимо коммуникативного контекста, служит способом передачи важной информации собеседнику, и научить робота распознавать эту информацию предстоит в будущем.

ЛИТЕРАТУРА

1. *Huggable™* [Электронный ресурс] / MIT Media Lab. Personal Robot Group. URL: <http://robotic.media.mit.edu/projects/robots/huggable/overview/overview.html> (дата обращения: 12.10.2012).
2. *Shrew whiskers inspire ground-breaking robot design* [Электронный ресурс] / Phys.org. URL: <http://phys.org/news/2012-01-shrew-whiskers-ground-breaking-robot.html> (дата обращения: 25.01.2013).
3. *Kismet* [Электронный ресурс] / MIT Artificial Intelligence Laboratory. Humanoid Robotics Group. URL: <http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html> (дата обращения: 12.10.2012).
4. *Leonardo* [Электронный ресурс] / MIT Media Lab. Personal Robot Group. URL: <http://robotic.media.mit.edu/projects/robots/leonardo/overview/overview.html> (дата обращения: 20.11.2012).
5. *MDS* [Электронный ресурс] / MIT Media Lab. Personal Robot Group. URL: <http://robotic.media.mit.edu/projects/robots/mds/overview/overview.html> (дата обращения: 20.11.2012).
6. *Breazeal C., Edsinger A., Fitzpatrick P., Scassellati B.* Active vision for sociable robots // IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans. 2002. Vol. 31, № 5. P. 443–453.
7. *Open N.I.* The standart framework for 3D design [Электронный ресурс]. URL: <http://www.openni.org> (дата обращения: 18.11.2012).
8. *Hema Swetha Koppula, Rudhir Gupta, Ashutosh Saxena.* Learning Human Activities and Object Affordances from RGB-D Videos [Электронный ресурс] / Cornell University Library. URL: <http://arxiv.org/pdf/1210.1207v1.pdf> (дата обращения: 4.10.2012).
9. *Spinello L., Arras K.O.* Leveraging RGB-D Data: Adaptive Fusion and Domain Adaptation for Object Detection // IEEE International Conference on Robotics and Automation (ICRA'12), St. Paul, USA, 2012 [Электронный ресурс] / University of Friburg. URL: <http://www.informatik.uni-freiburg.de/~spinello/spinelloICRA12.pdf> (дата обращения: 10.01.2012).
10. *Frederick Shic, Scassellati Brian.* A Behavioral Analysis of Computational Models of Visual Attention. // International Journal of Computer Vision. 2007. № 73(2). P. 159–177.
11. *Эко У.* Отсутствующая структура. Введение в семиологию. СПб., 2006. 548 с.
12. *Барт П.* Третий смысл // Строение фильма. Некоторые проблемы анализа произведений экрана. М., 1985. С. 175–188.

13. *Барт Р.* Camera Lucida. М.: Ad Marginem, 2004. 224 с.
14. *Лакан Ж.* «Я» в теории Фрейда и в технике психоанализа (Семинар Книга II (1954–1955)). М., 2009.
15. *Justin W. Hart and Brian Scassellati.* Mirror Perspective-Taking with a Humanoid Robot [Электронный ресурс] / Yale University. Social Robotics Lab. URL: <http://www.scazlab.com/justin/publications/HART-AAAI-12.pdf> (дата обращения: 27.10.2012).
16. *Киттлер Ф.* Мир символического – мир машины // Логос. 2010. №1 (74). С. 16–17.