A developmental algorithm for ocular–motor coordination

Chao, Fei; Lee, Mark Howard; Lee, Joseph J.

Published in:
Robotics and Autonomous Systems

DOI:
10.1016/j.robot.2009.08.002

Publication date:
2010

Citation for published version (APA):
Chao, F., Lee, M. H., & Lee, J. J. (2010). A developmental algorithm for ocular–motor coordination. *Robotics and Autonomous Systems*, *58*(3), 239-248. https://doi.org/10.1016/j.robot.2009.08.002

A Developmental Algorithm for Ocular-Motor Coordination

F. Chao[a], M. H. Lee[a], J. J. Lee[b]

[a]*Department of Computer Science*
Aberystwyth University, Wales, UK
[b]*John Radcliffe hospital*
Headington, Oxford, UK

Abstract

This paper presents a model of ocular-motor development, inspired by ideas and data from developmental psychology. The learning problem concerns the growth of the transform between image space and motor space necessary for the control of visual saccades. An implementation is used to produce experimental results and these are presented and discussed. The algorithm is simple, extremely fast, self calibrating, adaptive to change, and exhibits emergent stages of behaviour as learning progresses.

Key words: Developmental robotics, Sensory-motor coordination, Visual saccade learning

1. Introduction

Current research in robotics frequently draws inspiration from biology. By taking account of human and animal behaviour there is much to be learned about agents, autonomy and embedded cognition, and robotics has benefitted considerably by exploring new mechanisms and approaches.

There are three main sources of biological inspiration: structural biology, which deals with the anatomy and functioning of biological systems (e.g. neuroscience and endocrinology); evolution, which concerns growth and change across a population; and development, which addresses growth and change within the individual.

Of these three branches of biology both the structural and the evolutionary aspects have been intensively studied, as seen for example in robot controllers based on computational models of brain systems [30] and advances in evolutionary robotics [36], but only in recent years has developmental robotics blossomed as a research field; for a review see [28].

The aim of developmental robotics is to recognise processes of growth and change in human behaviour and build models that exhibit growth of competence and skill similar to that observed and reported by psychologists. It is notable that current robots still lack much of the adaptation, flexibility and learning seen in humans, and development may provide the key to further advances.

This paper describes a study inspired by the ability of the human eye to locate and rapidly move to targets to be examined. These eye movements are called saccades. We explore this ability in terms of a sensory-motor coordination context [37] and focus on very early infancy to produce developmental learning algorithms.

There have been many structural and computational models of rapid eye movements, e.g. [41, 19, 45, 13] for a review see [15], but we believe that our developmental approach [26] is able to produce a method with a unique set of features, namely: (i) it is not pre-wired but learns how to saccade, (ii) it learns very rapidly — much faster than current neural network based approaches, (iii) it continuously adapts to correct errors and accommodate any changes in the ocular-motor system, (iv) it does not use or require any calibration process or prior knowledge, and (iv) the generated behaviour displays distinct and qualitatively different stages which emerge during learning.

It is important to state that although our work is biologically-inspired, the aim is to create new mechanisms for controlling robots, not to directly contribute to the understanding of human behaviour. Hence, our models of biological systems often contain approximations and/or abstractions, and any similarity in behaviour does not necessarily signify that the same internal mechanisms are being used.

This paper is structured as follows: Section 2 presents an overview of the human ocular-motor system; Section 3 describes our visuo-motor coordination mapping model and the associated developmental learning algorithm; Section 4 explains an experimental implementation; Section 5 describes results from experiments; Section 6 discusses the findings and implications of this work, which is related to other research in Section 7; and finally, Section 8 gives a summary.

2. Human Visual Sensing and the Ocular-Motor System

The human visual system is not a passive receptor but must be actively directed at objects or features to be examined. During such examinations the eye is held onto the target; this is called fixation or foveation and the angular direction of the eyeball is known as the gaze.

2.1. The Human Eyeball

Each human eyeball is moved by six muscles; operating in pairs, they rotate the globe along orthogonal axes in three degrees of freedom. Two of the axes cover horizontal and vertical eye movements. The third muscle pair (rotation of the

retina) is used to correct the effects of torsion that can occur because rotations about intersecting spherical axes are non-commutative [50]. The eyes obey certain compensation laws and so do not display such effects; the reasons for this are still unknown but may be due to either special corrective neural circuits or the recently discovered muscle pulley structures [47]. In any case, this means torsional rotations can be ignored for our purposes.

The dynamics of the eyeball are very well behaved, primarily because there is almost no external loading (unlike all the other body parts) and the viscous and elastic properties are consistent over wide operating parameters. This allows control models to assume high repeatability of motor actions [48].

The ocular muscles are rich in spindle receptors that give high quality information about the stretch of the muscles which corresponds to the position of the eyeballs relative to the rest of the head. The stretch signals are effectively linearly related to angular rotation [23] and so high quality proprioception information on the direction of gaze is available.

2.2. The Retina

Unlike cameras, which have uniform sensor arrays with contiguous pixels usually arranged in a regular grid, the human retina is not uniform but consists of different sensor characteristics across its extent [11]. The periphery of the retina is a region containing low accuracy, monochrome sensing rods, which are very sensitive to change or movement. The central region contains colour sensing cones and consists of the macula or perifovea, which covers about 10 degrees of the visual field and the fovea, a region of densely packed cones which covers about two degrees and provides the greatest acuity and colour sensitivity.

There have been attempts to reproduce the sensing structure of the retina by transforming or mapping a retinal structure onto the uniform pixel arrays used in video cameras, but these have not been very successful. If a radial density function is used to increasingly space out the sensors towards the periphery then the fovea is not evenly spaced and a singularity can occur at the centre. Alternatively, if the fovea is evenly spaced but the periphery is not then a discontinuity is seen at the boundary. Balasuriya and Siebert argue that no analytical approach will produce a tessellation that gives uniform density coverage in the foveal and a space varying mapping in the periphery, while maintaining a consistent local structure [3]. They have produced tessellations that are good approximations to the cellular layout of the retina by using self-organising methods with random perturbations [3].

2.3. Eye Movement

Newborn infants are unable to track objects (smooth pursuit) or to discriminate motion direction, but they are able to perform saccadic eye movements. Saccades are very fast movements of the eyes that bring a stimulus from the periphery to the centre of the retina. Saccadic movements are generated by neural circuits in the brainstem and targets are selected by the superior colliculus which is a layered structure that includes visual and motor layers [34]. Vision is suppressed during a saccade, and saccades and fixations are mutually inhibited by the brainstem circuits [47].

Studies of human infants' saccades have shown that there are stereotyped age related changes in the way that their eyeballs move and that the more advanced movements coexist with earlier movements [42].

Saccades can move at speeds of up to 900° per second; which raises the question of how the brain knows when the eyeball has reached the target position. A feedback system can be ruled out because of the two candidate feedback signals, retinal signals would not be processed in time [48] and proprioception has been shown to be not necessary for accurate saccades [16]. The rate of firing of motor neurons that drive the eye muscles are linearly related to eye position [48] and this has formed the basis of a feedforward mechanism for eye position control. This widely accepted idea is known as the "corollary discharge" or "efferent copy" model [16] in which motor values are taken as reliable indicators of eye position. However a feedforward model is subject to local errors or drift and proprioceptive receptors and optokinetic functions are assumed to have a major role in stabilisation and holding the gaze during fixations [6].

Thus the visual location of the target for a saccade and the motor values that bring the target to the centre must be correlated and this means there must be a close coordination between the image space of the retina and the motor space of the ocular muscles. Such sensory-motor coordination must be maintained as an internal program or function which must be either innate or learned. For this to occur innately would require detailed prior knowledge of the muscular system and the optical characteristics of each particular eyeball, and it seems almost inconceivable that this would happen in the newborn infant.

The issue of innate versus learned behaviour has been debated between psychological empiricists and nativists over many decades. Because newborn infants can produce saccades it is generally assumed that this is an innate competency, but we demonstrate here that very rapid learning is a feasible alternative possibility.

3. A Model of Ocular-Motor Coordination Learning

Our robot system has only one eye and the head is fixed in space. This simplifies the system compared to the human, and we note that infants do not integrate head movements until around two months after birth [41].

The main control issue for the ocular-motor system is a sensory-motor coordination problem: what are the necessary motor variables to drive the eye to move the foveal area to a specific sensed peripheral region? This actually contains two coordination problems, first the retinotopic image space must be related to the eyeball gaze space, and then the target eyeball location must be translated into motor commands. However, accepting that saccade action must involve feedforward mechanisms as described in section 2.3 and given that gaze angle is linearly related to motor firing rates [48] we can assume that the gaze/motor relationship is direct and linear (i.e. an element

transform). This means we do not need to model the gaze/motor relation and can rely on the motor plant to perform ballistic movements from a given current gaze location to a desired target position [31]. Thus, a gaze value is always equivalent to a motor command and in the following we will refer to either gaze points or motor data according to context. Furthermore, this target-driven control also means that the dynamical aspects of saccades, that is, parameters such as velocity, gain and duration which have been extensively examined [12], are not relevant to our model. Should the gaze/motor relation not be trivial then it could be learned and we note that the motor systems of the eye are exercised in the womb [4] and hence motor activation signals could be correlated prenatally with eyeball position via the proprioceptive signals from the muscle spindle receptors.

To summarise, the problem we address here is the growth of the transform between image space and eye gaze space. Our assumptions are that (a) this may not be a simple or linear relationship and (b) the position of the eyeball can not be related to image data until after birth (as vision is ineffective in the womb [44]), and therefore learning should start from zero prior knowledge.

In our work we have used a mapping technique to model sensory-motor relationships [25] and we adopt this method here. Each channel of either sensory or motor information is provided with a two-dimensional map consisting of many overlapping elements. These elements, known as fields, represent patches of receptive area on which stimuli fall. In our models we use sheets of fields that are circular and overlapping. All stimuli that land within a field are represented by the coordinates of the field centre; thus, fields can be thought of as a tolerance or resolution limit. Our system has image data as

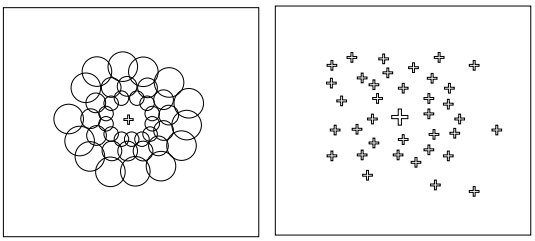

Figure 1: The Ocular-Motor Map Layers

the sensory input and a two-degree motor system for moving the image. Thus, two map layers are needed and these are illustrated in figure 1. The left layer is a visual sensory map which uses polar coordinates because a polar mapping is the natural relation between central and peripheral regions on the retina [46]. The layer on the right in figure 1 is the associated motor drive layer; this is a motor map in two degrees of freedom and encodes the horizontal (left-right), and vertical (up-down) eye movements. As correspondences between fields on different layers are discovered by experience so they become directly linked. That is, when a movement causes an accurate shift of the fovea to a periphery stimulus, then the sensory field (giving the stimulus location) is explicitly coupled to the motor field (giving the motor variables that produce the change). By this means, the sensory-motor relations for accurate saccades are discovered and learned.

Following the human retinal structure, we designed the field density to be higher in the central area than the periphery. This was achieved by generation rules that allow field spacing and radius to vary with distance from the centre. Fields are assigned on a grid consisting of two sets of 20 radial lines at $360^\circ/20 = 18^\circ$ separation, with the two sets offset by 9°. The distance of each field centre from the origin is calculated according to the rule: $d_i = 1333(\alpha^i - 1)$ for $i = 1 \ldots 20$, where $\alpha = 1.015$ is a spacing factor, and each consecutive ring of fields is placed on alternating sets of radial lines. This gives a close packing similar to the hexagonal grid used to efficiently pack circles into rectangular areas. Figure 2 shows this structure in detail. The radius of each field was related to distance from the centre by a rule that gives significant field overlap: $radius_i = \pi\beta d_i$, for $i = 1 \ldots 20, \beta = 0.06$.

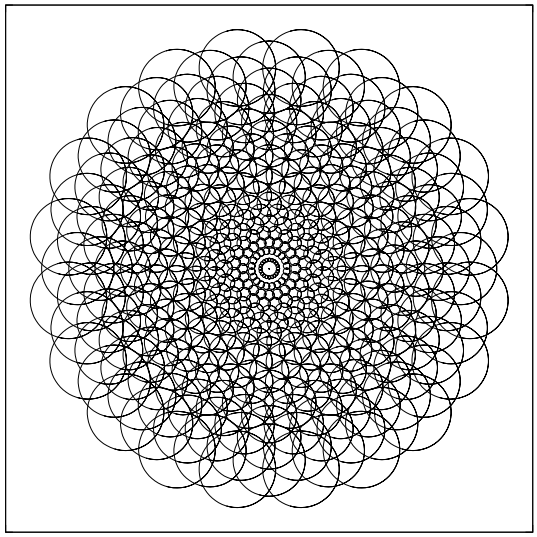

Figure 2: The grid of fields for the image map

It is important to note that the grid in Figure 2 is not an example of a visual sensory map but only illustrates the potential locations of where field centres may be placed. Fields are only created on the visual map when a stimulus lands in an area that is not covered by any fields. To produce a complete covering requires less fields than in the grid and so we expect only a proportion of fields seen in this grid structure will be used in the image map.

This design means that longer saccades may be more inaccurate than shorter ones. However, there is some evidence that the more extreme peripheral saccades are more inaccurate in studies on infants [22] and adults [27]. The motor coordinate system is Cartesian, as in Figure 1, because the eye muscles are independent and orthogonal,

These two mapping layers are initially empty and are not prewired or pre-structured for any specific spatial system. Fields are created when new sensory-motor values are to be recorded and the maps become populated according to the pattern of experiential events.

3.1. The Developmental Learning Algorithm

Previous research shows that two week-old infants scan geometric figures rather randomly, while fourteen week-old infants direct their saccades to stimulus contours more consistently [7]. It seems that saccadic eye movements are refined over a period. However, Butko and colleagues proposed a rapid learning hypothesis which argues that very fast learning might occur just after birth [8]. We suggest such fast learning for eye saccades might be obtained by the following process: when an infant senses an object appearing in her field of vision, the infant's brain is stimulated to try to move her eyeballs to fixate the object. However, because all visual data is novel to a newborn infant we assume there is no prior coordination of retinal image space with motor acts. Hence, the appropriate motor values are not known and so the infant's brain may generate spontaneous (random) motor values in an attempt to move towards the object. When, eventually, the infant's eyes fixate on the object, then the brain can record the parameters of the successful experience (initial peripheral location, final angle of gaze) for future reference.

An autonomous learning algorithm can be developed to reflect the above learning process and this is summarised as pseudo code in Figure 3.

```
For each session
    If stimulus in peripheral vision at θ, γ
        Access the ocular-motor map
        If a covering field exists:
            Use motor values for this field
        Else
            Record the stimulus position,
            make a spontaneous motor move
            If the stimulus is within the fovea:
                Generate a new field,
                enter the stimulus location and
                the associated motor values
            Else
                Repeat
            End If
        End If
    Else
        Do not move
    End If
Iterate a new session
```

Figure 3: An Elementary Algorithm

This simple baseline algorithm is dramatically improved by the addition of the following two modifications.

3.1.1. Nearest Field Selection

Suppose that the ocular-motor map has not yet generated any fields that cover the current periphery stimulus location, let this be (θ, γ). The *nearest field* to the stimulus can then be selected as an approximation to the target. For this, the following nearest selection procedure was designed: first, an angular tolerance is set to select the fields which have a similar angle with the target field (θ), this tolerance is thus defined: $\theta \pm \delta_1$. Then, a distance tolerance is set to select the fields nearest (radially) to the target field from amongst the candidate fields in the above set. The distance tolerance is defined as: $\gamma \pm \delta_2$ pixels. The angular parameter is given precedence over distance because, in polar coordinates, the angular coordinate alone is sufficient to determine the trajectory to the origin. From this we can obtain a set of fields which fall within the (broad) neighborhood of the stimulus, and the following formula

$$\text{MIN}(\sqrt{(\gamma - \gamma_x)^2 + (\theta - \theta_x)^2})$$

is used to choose the nearest field from this collection, where γ_x and θ_x, for $x = 1 \ldots n$ are the fields in the collection. This is summarised as pseudo code in Figure 4.

```
If no fields exist for location θ, γ:
    a. For each field, f_x ∈ Fields
        If θ − δ1 < f_x(θ) < θ + δ1
            Candidates = Candidates ∪ {f_x}
    b. For each field, f_x ∈ Candidates
        If γ − δ2 > f_x(γ) or f_x(γ) > γ + δ2
            Candidates = Candidates − {f_x}
    c. Apply the MIN formula to Candidates
       to find the nearest field to θ, γ.
```

Figure 4: The Nearest-Field Selection Algorithm

Note that although the variables in the MIN calculation use different units their ranges are compatible for our purposes (being 0-360 degrees and 0-300 pixels). In the experiments (see Section 5), δ_1 is set to 15° and δ_2 is set to 10 pixels.

3.1.2. Vector Field Generation

In the basic algorithm given in Figure 3, a new field cannot be generated until the camera has fixated an object at the target location, and this process typically takes a long time because most spontaneous moves will not result in a target fixation. However, we note that there is a change in the location of the stimulus in the image after *each* movement. A vector can be produced from this change by:

$$\vec{V} = Position_{old} \text{ - } Position_{new}$$

where $Position_{old}$ denotes the object position before movement and $Position_{new}$ the object position afterwards. This vector represents a movement shift of the image produced by the related motor action. Consequently, the vector can be used to access a field in the image layer together with its corresponding motor values on the motor layer. In so doing, a new field can be generated after each spontaneous movement. This idea is related to the Hebbian learning model [33].

During very early learning many spontaneous movements will be needed until a fixation is achieved and by using the movement vector idea each fixation can generate many vectors. At any time, the current vector will be a sum of the previous vectors, thus:

$$\vec{V_s} = \sum_{i=1}^{n} \vec{V_i}$$

and the corresponding motor values, being linear, can also be produced by summation:

$$M^s_{(p)} = \sum_{i=1}^{n} M_{i(p)},\ M^s_{(t)} = \sum_{i=1}^{n} M_{i(t)}$$

where p and t are the independent eye movement axes.

This is an incremental and cumulative system, in that the resultant vectors can be built up over a series of actions by a simple recurrence relation:

$$\vec{V}_{sum}(t+1) = \vec{V}_{sum}(t) + \vec{V}_i(t+1)$$

4. System Implementation

Figure 5: The Pan and Tilt "Eye" System

Our laboratory robot incorporates a motorised camera system that acts as an "eye". Figure 5 shows the hardware components consisting of a video camera mounted on a pan-and-tilt head.

4.1. The Motor Subsystem

The motor system is implemented by a motorised pan-and-tilt device which provides two degrees of freedom. The pan motor can drive the video camera to rotate about an axis that translates the image in one direction, and the tilt motor can drive rotation about an orthogonal axis, giving image translation at 90 degrees. Combined movements of pan and tilt motors cause motion along an oblique axis. The pan/tilt device can effectively execute saccade type actions based on supplied motor values from the learning algorithm. Each motor is independent and has a value (M_p for pan and M_t for tilt) which represents the relative distance to be moved in each degree-of-freedom.

4.2. The Sensor Subsystem

The camera captures workspace images and image processing software is used to implement two sensors: a periphery sensor and a centre or foveal sensor. The periphery sensor detects new objects or object changes in the visual periphery area and also the positions of any such changes (encoded in polar coordinates). The centre sensor detects whether any objects (i.e colour blobs) are in the central (foveal) region of the visual field. Figure 6 shows the workspace which is a white table, with green objects. This setup was arranged to simplify the image processing task, especially object detection.

The camera capture rate is one frame per second. A circular area, of radius 20 pixels, in the centre of the image is defined as the foveal region. If the centroid of an object is in this central area, it is considered that the object is fixated and saccades are inhibited; otherwise the system is not fixated. Each object is represented by a group of green pixels clustering together in the captured image. The position of the centroid of the pixels is used as the location of the object. The image processing program compares the currently captured image against the stored previous image and, if the number or the position of any central pixels within these two images differs markedly, the object is considered to have changed and the new location of the centroid is encoded in polar coordinates.

Note that an object "change" here signals one of the following three situations, (i) an object is moved to a new location in the workspace; (ii) an object is removed from the workspace; and (iii) a new object is placed in the workspace. Of course, moving the camera would also cause a change in an object on the retina but, as for the human eye, image processing is only performed during fixations, not saccades. When one object disappears and another reappears, young infants will often approach the new stimulus with a series of saccades rather than only one [2]. However, the current version of the algorithm merely uses single object appearance as the stimulus. As an experimental technique we did not manually move the object but simply moved the camera to a random location when a "new" object at a different location was required at the start of a run.

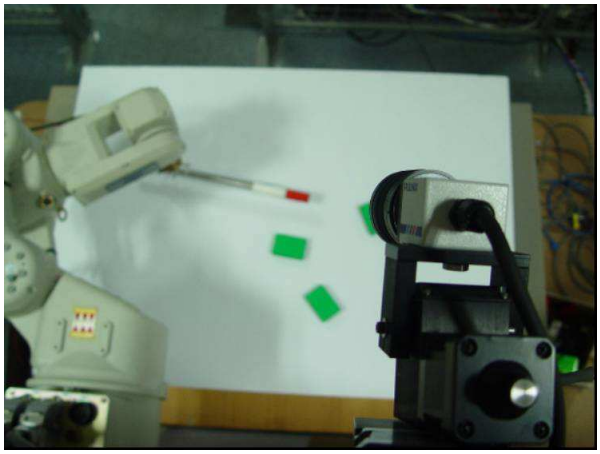

Figure 6: A View of the Workspace

5. Experimental Results

The experiments are designed to investigate the model described in section 3 The experimental procedure is ordered as follows: an object is placed within the camera's field of view, then the developmental learning algorithm drives the camera until the object is fixated; after fixation, the object is moved to a new position (still within the camera's view) and this process iterates. During this procedure, no people or other agents are

involved except for moving the object's position. Through repeated experiments, it is hoped that all or most of the locations in visual space will have been covered so that most possible fields in the ocular-motor map will have been created.

5.1. Observations

From a large number of experiments carried out, we observed that this system's behavior can be described as falling into three stages: (1) at the beginning of a new ocular-motor map, (2) after a few fields have been generated, and (3) after most fields have been created.

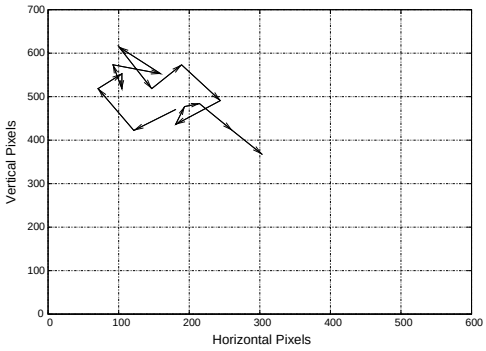

Figure 7: First Stage Traces, (image centre at 300, 350).

Figure 7 shows the traces of movement on the visual image at the beginning of a new ocular-motor map. Actually the object is static during the experiment, but the camera is moving; hence, in the image, the object appears motile. During this early stage, because the new ocular-motor map is blank or extremely sparse, there is no experience available (in terms of nearby fields) and thus most movements are simply spontaneous. In the example in figure 7 there are fifteen traces before fixation is achieved.

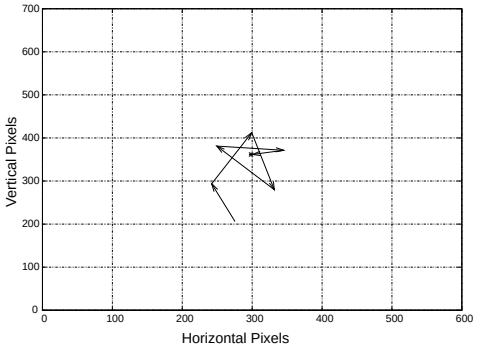

Figure 8: Second Stage Traces

When a moderate number of fields have been generated, it is still difficult to find an exact corresponding field for the stimulus, but the nearest-field algorithm usually finds a nearby field. Figure 8 illustrates the process of this second stage: large spontaneous movements do not happen any more, and the movement traces tend towards the image centre.

At the third stage, because most fields have been generated, the learning algorithm is able to find the correct corresponding field (and thus the associated motor values) each time, the camera movement is much simpler, usually consisting of one movement and fixating the object directly. Figure 9, comprising six experimental results in one plot, shows the traces as radial movements, from periphery to centre; note that one of the plots required two saccades to reach the target, indicating that map learning was not yet fully complete.

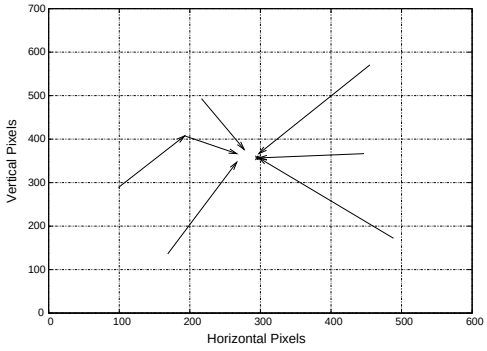

Figure 9: Third Stage Traces

Figure 10 illustrates the outcome of the set of experiments: the upper figure presents the sensory layer and the lower figure the motor layer. It can be seen that much of the image map has been covered with fields, in this run a total of 94 fields were produced. The fields in the sensory layer are plotted in polar coordinates and marked by numeric labels, which give correspondence with the motor (gaze) values.

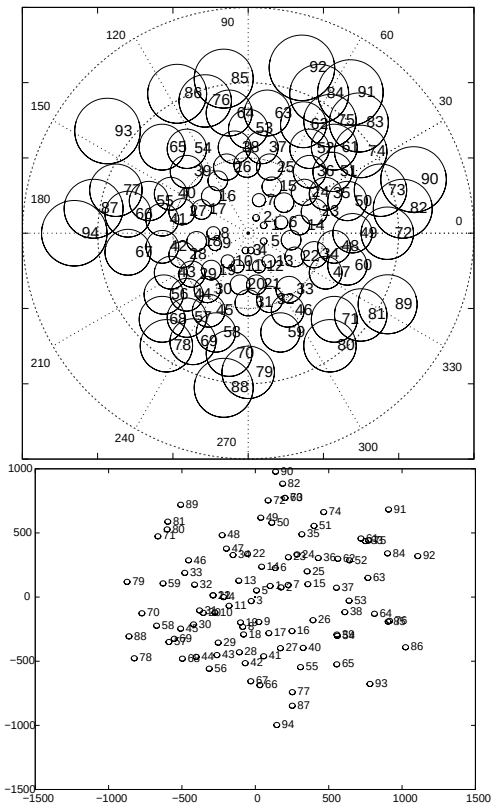

Figure 10: The Ocular-Motor Map at 94 fields. In the sensory layer radial contours are drawn at 100, 200 and 300 pixels. The motor values represent relative displacements.

During the experiments each movement was recorded and flagged as one of three types: spontaneous, (no suitable field exists); using a neighbouring field; or direct saccade (stimulus covering field found). Figure 11 is a cumulative plot that shows the mix of movement types over time; three runs for each type are shown to illustrate the variation in the process:

• The number of spontaneous movements (type A) dominates

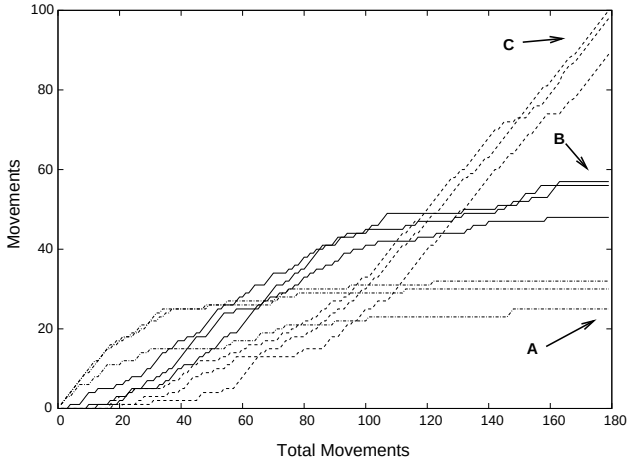

Figure 11: The Three Types of Movement. Three experimental results are plotted in this figure, each line-style stands for a type of movements.

during the first thirty movements, however, this type of movement occurs very little from then on.

• Movements using nearest neighbour fields (type B) do not exist at the beginning, but this type of movement increases sharply after that, and then after a period of growth, around 90, the use of nearest fields becomes less frequent.

• Direct, accurate movements using the correct corresponding fields (type C) do not occur at all during the first eighteen movements, however at the end of the experiments these have the fastest rate of increase, until finally only these single saccades exist.

In order to illustrate where the fast learning occurs, Figure 12 shows the rate of new field generation over an entire experiment. As can be seen, the field generation rate produced by the developmental learning algorithm is very fast for the first 80 movements (at the rate of one field for every 1.27 moves), then the rate decreases, and finally, field creation becomes very rare.

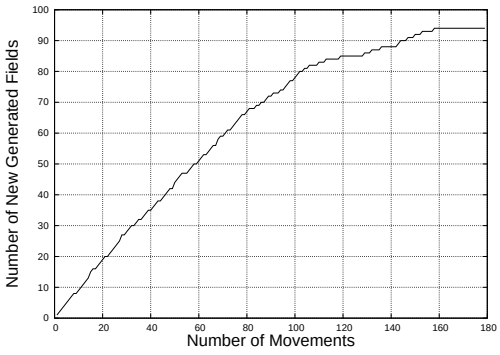

Figure 12: The Rate of New Field Generation.

Another illustration of the learning process is seen in Figure 13, taken from another, different run of experiments. This data covers 88 movements in total, separated out into saccades per individual fixation. This shows how the number of saccades per fixation falls away very rapidly, the reason being that even a sparse covering of fields aids convergence because a near neighbour can usually be found.

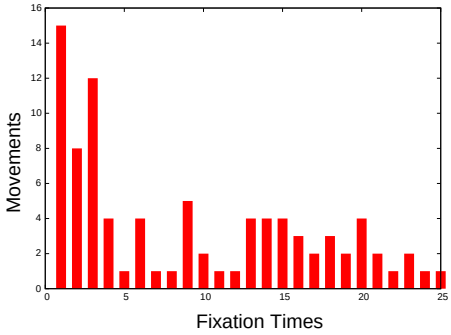

Figure 13: The Decline in Saccades per Fixation.

It is important to note that the emergence of the observed behavioural stages are not initiated by any switching or thresholds. Indeed, at any point the system might revert to an earlier behaviour type, as at any time a new field might need to be introduced. Eventually the early behaviours will be extinguished but as this is an asymptotic process there is always a finite possibility of regression.

6. Discussion

It has been suggested that the preference of newborns to orient towards faces is not innate, as generally believed, but could be learned very rapidly, even in the first six minutes of life [8]. Our robot model has demonstrated that the fundamental process of visual saccading to a peripheral stimulus could also be learned rather than innate. If such rapid learning does occur it would be quite difficult to detect but it could be very significant evidence for the empiricist stance [49].

From a robotics design viewpoint is is also difficult to see how accurate saccades could be built as an innate function. In order to coordinate points on a camera image plane with objects in the 3D world it is necessary to analyse the coordinate geometry of the imaging process and provide some form of calibration procedure. It seems very unlikely that such concrete and specific information could be transmitted by innate means. It is important to note that our model does not need any such information and is self-calibrating; indeed the learning model is essentially performing a kind of continuous calibration learning.

There is a large literature on eye movements and saccades but very little is relevant to the first few hours and days after birth. However, we find considerable support for our model which seems entirely compatible with current knowledge. For example, it is known that infants execute smaller steps than adults

with reports of "sometimes making as many as four or five consecutive saccades to reach the peripheral target" [2], and "larger step sizes were used to localize more distant targets" [2].

As in the robot, the average number of saccades an infant uses to fixate an object onto its fovea reduces with time. Roucoux et al. found that infants fixed targets onto the fovea with successive small saccades even though they were capable of larger saccades [42]. They also found that the more eccentric the target the higher the average number of saccades, and for targets at all angles the average number of saccades declined with age. This is similar to the robot model, in which the number of saccades decreases over time.

Roucoux et al. describe two patterns of movement that result in the image falling on the fovea, one with predominant head movements and small saccades and the other being the more adult pattern of larger saccades predominating [42]. They draw a parallel with two types of vertebrates, those with a fovea and those without and designate the head dominant movements as "afoveate" and the more saccadic as "foveate". They describe how, as infants mature, the proportion of "foveate" movements increases and the range of angles from the fovea that are covered by "foveate" moves or saccades increases. For objects nearer the centre of vision the adult-like pattern was reached earlier than for more eccentric targets. At five weeks targets of 45 degrees were not fixated at all, by 8 weeks they took 3 saccades on average and by adulthood this was almost one. These observations are consistent with staged learning behaviour and, allowing for the complications caused by head movements, they are similar to the pattern seen in our robot model.

Hainline and colleagues examined saccade peak velocity, amplitude and duration in infants and compared them to adults [17]. Their sample of 64 infants of ages 14 to 151 days produced a significant proportion of mature saccades that were comparable to adults. This was confirmed by [14] who also showed that infant saccades may even be faster than adults. The mix of infant and adult movement types reported by such authors can be interpreted as part of a learning process and the decline of the early patterns with age reflects the increasing dominance of the more efficient saccades. This pattern is seen in the robot model, where mature and immature movement patterns (from the 3 stages) coexist until the system has fully learnt the relationship between its motor and visual maps.

An interesting hypothesis [18] is that the commonly observed undershoot in saccades may be an optimum strategy to minimise the total flight-time, because the total flight-time is less with corrective saccades that undershoot as compared with those that overshoot. This effect also occurs in our method — fields near the foveal are predominant among the first to develop because most moves end in such a field, and so when a neighbour is selected it is more likely to be on the near side than the far side of the target. We analysed the data for a run of fixations and found that undershoot occurred in 75% of the cases.

Regarding the robustness of the method, we notice that the motor values are, in fact, not absolute gaze positions but define shifts in gaze relative to the current position. This means a periphery stimulus at point P on the image might cause a saccade from the current gaze location G_1 to new location G_2 but then another stimulus occurring at exactly the same point, P, on the image will drive the system to G_3. This means the gaze space must be linear — an image shift must always produce the same *change* in gaze for all gaze locations. As mentioned before, if this was not the case then a gaze/motor mapping could be used to hold the corrections necessary to linearise the gaze space. On the other hand, there is no requirement for the image space to be linear and the system effectively learns image distortions. This means that although the summation of motor values during multiple movement learning will produce a correct result, the summation of vectors on the image is likely to produce errors. Nevertheless, the fields represent a "zone of tolerance" and small errors will often be accommodated within fields. In our experiments we have found that these were indeed accommodated in maps for typical camera/lens combinations, as the intermediate vector generation method was successful in speeding learning. In the case of more difficult mappings there are two options: either field generation for intermediate vectors is switched off, notice that this will only slow down learning; or it may be more effective to continue creating fields wherever possible and allow later corrections to be made. This question of rapid population and correction or slower more accurate growth requires more investigation as different conditions may apply for different tasks. We also notice that the total gaze space will be larger than the image and so it is possible for the spontaneous movements to shift the target out of view (off image), as happens occasionally in our experiments. This does not cause any problems as the motor values continue to be accumulated and the final motor summation is still the appropriate value for the stimulus target field. During such events the intermediate vectors can not be utilised but when the stimulus returns to the image the process continues as before.

Regarding plasticity, consider how the algorithm would perform with a fully populated mapping and then the image is subjected to some fixed distortion through optical or physical disturbance. A peripheral stimulus will still be covered by a field but the motor movement will be incorrect (by an amount depending on the degree of distortion) and the foveal region is likely to be missed. A second, corrective saccade will then be triggered by the new field location and this is likely to reach the fovea as it will be much nearer. Thus, we would expect a small number of corrective saccades to be generated and the final motor values, $M^s_{(p)}$ and $M^s_{(t)}$, can be inserted to replace the previous values, (effected by a trivial adjustment to the algorithm in Figure 3). Thus the system will adapt to changes as they are experienced. If only part of the image is distorted then only part of the mapping will be relearned, but even in the worst case, the time taken will be no longer than learning the original map, and this can be done on-line and during use. Of course, if the distortion is of a warping nature then corrections will be local and the process will be well behaved. On the other hand, a gross change, such as a total inversion in a reflecting mirror would require a completely new map to be learned, while a change of lens focal length (zooming), would be an intermediate change with the center unaltered and the periphery notably shifted.

Considering the accuracy of the system, the model can easily match reported infant accuracy [22, 20]. In the mapping, the

average error in saccading to a given image location is $0.35R$ for a field of radius R, for double overlapping fields — this is always within the tolerance provided by the foveal region of 20 pixels. But note that using a simple linear field function (e.g. stimulus distance from field centre) a multilateration operation performed on a small number of local fields will deliver much higher accuracy if needed. Full details of noise analysis and the effects of overlapping fields on accuracy requires a further paper.

7. Related Work

Many robotics projects have involved ocular-motor coordination problems, and indeed most robots with hand and eye systems need to deal with tasks such as visually directed gaze control, visually guided reaching, and other human inspired sensory-motor behaviour. However, the ability to saccade to visual stimuli of interest has often been programmed into many systems rather than learned. Maybe the image/gaze relation has often been considered innate and therefore can be engineered as a fixed function [10]. Even when saccades have been learned they have often not been very closely aligned with existing psychological data and knowledge. For example, [24] addressed the problem of driving moveable visual sensors to locate static objects. The emphasis was on on topographic mappings and artificial neural networks, but the neural controller needed 100,000 trials during training. [21] produced a very novel artificial evolutionary method that simulated genes and regulators to create a neural network that learns to track objects on an image. This required 30,000 training iterations and the resulting image tracking function was more of a retinal flow field that produced somewhat distorted paths to the fovea rather than direct saccades from any image point. A strong developmental approach to visually guided reaching has been described by [32]. Motor synergies or primitives were used to provide a motor/motor correlation learning system for the hand and eye components. A very similar approach to our own is seen in [1] where visual tracking needs to be learned as part of a sensory-motor approach to imitation research. The visual image was mapped onto a robot hand map but small, local neural networks were incorporated into *every* field point. The system learned to coordinate a 2 degree-of-freedom image with a 3 degree-of-freedom robot arm but required 8,000 random movements.

In many such robotic models of saccading and ocular-motor coordination, we find that the learning times reported are usually orders of magnitude greater than for our system. Connectionist methods have been widely employed, and although techniques such as radial basis functions have similarities with our mappings, it seems that the extensive training regimes required are not very compatible with experience of human development. Also the number of neurons involved with these methods can scale up exponentially [38]. The system of [29] is also similar in motivation to our own but uses engineering solutions where psychological methods would produce more efficient and flexible results. For example, saccade maps were learned but these were primed with a linear grid of 10x10 gaze locations, and then learning was used to adjust the errors to the true map. But even with this prior knowledge, each location required 20 trials, thus giving 2000 learning trials [43]. In comparison our method produces a nearly complete mapping by 200 trials. Also, before learning, a calibration routine was used, which gave undefined prior information to the system. A hand/eye mapping was also learned by this system but all the mappings described had to be learned as separate functions in each direction (i.e. each map and its inverse), whereas the links in our mappings are bidirectional — the image/gaze map can be read in reverse to find the expected location on the retina that corresponds to a planned gaze shift.

The use of random movements or "motor babbling" has been valued by a number of researchers. Harris argues strongly that sophisticated control theory is inappropriate for modelling saccade control because this is an un-referenced control problem (i.e. no error reference is available) [18]. This means exploration of the problem space is necessary and "randomness (variability) of activity reflects an active process for exploring... rather than being simply neural noise" [18]. It is important to recognise that "random" action can be much more than "trial and error" learning, which is a form of blind search. Appropriate spontaneous action is an information gathering activity and can be a rich source of data for learning forward models.

Although this work has no direct link with neuroscience data or models, the resulting system is not incompatible with a neural interpretation and it would be feasible to implement the algorithm as a version based only on artificial neural network techniques. Such a version could take advantage of parallelism, in nearest field selection for example, to provide even faster real-time operation.

8. Conclusions

The experimental implementation described here demonstrates that our developmental approach is able to produce a method that: learns very rapidly — much faster than current neural network based approaches; does not use or require any calibration process or prior knowledge; continuously adapts to correct errors and accommodate any changes in the ocular-motor system; and displays distinct and qualitatively different stages in its behaviour which emerge during learning.

Our model draws on the relevant literature and our resulting experimental system is extremely fast, incremental and cumulative in its learning; all desirable characteristics for real-time autonomous agents. This relates to human infant learning and adaptation which has often been observed to be very fast [40]. The simplicity of the method is important in this regard, as the requirement for fast performance, in *both* saccading and learning, rules out complex computations and the speed of neural processes limits the number of steps that can be involved [35, 5].

Developmental psychology recognises a key characteristic of animal development: the sequencing of development phases where some competencies always precede others. These regularities are known as stages and are believed to be the basis of development processes that underpin the gradual consolidation of control, coordination and competence [39]. The challenge

from this viewpoint is in finding effective algorithms that support progressive and qualitative growth in behavioural competence without requiring significant structural change. Our results show three distinct stages of behaviour emerging from a single process; this shows how qualitative change in behaviour may occur without structural change, but by the consolidation of experience. This concept of staged growth under constraints may provides a valuable method for use in many sensory-motor learning applications [26].

Regarding potential applications, our algorithm provides automatic fixation of stimuli points in a visual field and thus would be valuable in moving camera applications such as surveillance, monitoring, undersea, and rescue situations, particularly when the system is mobile, temporary or vehicle mounted. The avoidance of any calibration, set-up, or training periods is a great advantage. Many existing methods deem it necessary to establish exact correspondences between video images and the 3D sensed environment and this requires elaborate computations of intrinsic and extrinsic camera geometry, often with the use of calibration objects [51]. These methods can take up to 30 minutes for the calibration process [9]. Our simpler approach does not need any camera parameters and yet can handle non-linear image distortions. A patent application is currently in progress.

Further work can build on this model for the growth of further behaviours, including corrective saccades, smooth pursuit, head integration and gaze analysis. There exists a rich source of psychological data that can provide guidance for building effective learning algorithms that advance robotic applications and there remains much to be done in implementing working developmental algorithms in autonomous agents.

9. Acknowledgements

This work was mainly supported by UK EPSRC grand GR/R69679/01. The first author was also supported by an ORS award.

References

[1] Andry, P., Gaussier, P., Nadel, J., Hirsbrunner, B., 2004. Learning Invariant Sensorimotor Behaviors: A Developmental Approach to Imitation Mechanisms. Adaptive Behavior 12 (2), 117.

[2] Aslin, R., Salapatek, P., 1975. Saccadic localization of visual targets by the very young human infant. Perception and Psychophysics 17 (3), 293–302.

[3] Balasuriya, S., Siebert, P., 2005. A biologically inspired computational vision front-end based on a self-organised pseudorandomly tessellated artificial retina. In: Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on. Vol. 5.

[4] Birnholz, J., 1981. The development of human fetal eye movement patterns. Science 213 (4508), 679.

[5] Braitenberg, V., Schüz, A., 1998. Cortex: Statistics and Geometry of Neuronal Connectivity. Springer.

[6] Bridgeman, B., Stark, L., 1991. Ocular proprioception and efference copy in registering visual direction. Vision Research 31 (11), 1903–13.

[7] Bronson, G., 1990. Changes in infants' visual scanning across the 2- to 14-week age period. J Exp Child Psychol. 49 (1), 101—25.

[8] Butko, N. J., Fasel, I. R., Movellan, J. R., 2006. Learning about humans during the first 6 minutes of life. In: Fifth International Conference on Development and Learning. Bloomington, Indiana, USA.

[9] Chen, H., Matsumoto, K., Ota, J., Arai, T., 2007. Self-calibration of environmental camera for mobile robot navigation. Robotics and Autonomous Systems 55 (3), 177–190.

[10] Coelho, J., Piater, J., Grupen, R., 2001. Developing haptic and visual perceptual categories for reaching and grasping with a humanoid robot. Robotics and Autonomous Systems 37 (2-3), 195–218.

[11] Curcio, C., Sloan, K., Kalina, R., Hendrikson, A., 1990. Human photoreceptor topography. Journal of comparative neurology(1911) 292 (4), 497–523.

[12] Deubel, H., Wolf, W., Hauske, G., 1986. Adaptive gain control of saccadic eye movements. Hum Neurobiol 5 (4), 245–53.

[13] Gancarz, G., Grossberg, S., 1998. A neural model of the saccade generator in the reticular formation. Neural Networks 11 (7-8), 1159–1174.

[14] Garbutt, S., Harwood, M., Harris, C., 2006. Infant saccades are not slow. Developmental Medicine and Child Neurology 48 (08), 662–667.

[15] Girard, B., Berthoz, A., 2005. From brainstem to cortex: computational models of saccade generation circuitry. Progress in Neurobiology 77 (4), 215—51.

[16] Guthrie, B., Porter, J., Sparks, D., 1983. Corollary discharge provides accurate eye position information to the oculomotor system. Science 221 (4616), 1193–1195.

[17] Hainline, L., Turkel, J., Abramov, I., Lemerise, E., Harris, C. M., 1984. Characteristics of saccades in human infants. Vision Research 24 (12), 1771–1780.

[18] Harris, C., 1998. On the optimal control of behaviour: a stochastic perspective. J Neurosci Methods 83 (1), 73–88.

[19] Harris, C., Berry, D., 2006. A developmental model of infantile nystagmus. In: Seminars in Ophthalmology. Vol. 21. Informa Healthcare, pp. 63–69.

[20] Harris, C. M., Jacobs, M., Shawkat, F., Taylor, D., 1993. The development of saccadic accuracy in the first seven months. Clinical vision sciences 8 (1), 85–96.

[21] Hotz, P., Gomez, G., Pfeifer, R., 2003. Evolving the morphology of a neural network for controlling a foveating retina-and its test on a real robot. In: Artificial Life VIII-8th International Conference on the Simulation andSynthesis of Living Systems. Vol. 2003.

[22] Kowler, E., Blaser, E., 1995. The accuracy and precision of saccades to small and large targets. Vision Res. 35 (12), 1741—54.

[23] Krauzlis, R., 2005. The Control of Voluntary Eye Movements: New Perspectives. The Neuroscientist 11 (2), 124.

[24] Kuperstein, M., 1991. INFANT neural controller for adaptive sensory-motor coordination. Neural networks 4 (2), 131–145.

[25] Lee, M. H., Meng, Q., Chao, F., 2006. A content-neutral approach for sensory-motor learning in developmental robotics. In: Proceedings of the 6th International Workshop on Epigenetic Robotics. Paris, pp. 55—62.

[26] Lee, M. H., Meng, Q., Chao, F., 2007. Staged competence learning in developmental robotics. Adaptive Behaviour 15 (3), 241–255.

[27] Lewis, A., Garcia, R., Zhaoping, L., 2003. The distribution of visual objects on the retina: connecting eye movements and cone distributions. J Vision 3 (11), 893—905.

[28] Lungarella, M., Metta, G., Pfeifer, R., Sandini, G., 2003. Developmental robotics: a survey. Connection Science 15 (4), 151–190.

[29] Marjanovic, M., Scassellati, B., Williamson, M., 1996. Self-Taught Visually Guided Pointing for a Humanoid Robot. In: From Animals to Animats 4: Proc. Fourth Int l Conf. Simulation of Adaptive Behavior. pp. 35–44.

[30] Mataric, M., 1991. Navigating with a rat brain: A neurobiologically-inspired model for robot spatial representation. From Animals to Animats, MIT Press, Cambridge, MA.

[31] Mays, L., Sparks, D., 1980. Saccades are spatially, not retinocentrically, coded. Science 208 (4448), 1163–1165.

[32] Metta, G., Sandini, G., Konczak, J., 1999. A developmental approach to visually-guided reaching in artificial systems. Neural Networks 12 (10), 1413–1427.

[33] Munakata, Y., Pfaffly, J., 2004. Hebbian learning and development. Developmental Science 7 (2), 141—148.

[34] Munoz, D., Pelisson, D., Guitton, D., 1991. Movement of neural activity on the superior colliculus motor map during gaze shifts. Science 251 (4999), 1358–1360.

[35] Nagarajan, N., Stevens, C., 2008. How does the speed of thought compare for brains and digital computers? Current Biology 18 (17), 756–758.

[36] Nolfi, S., Floreano, F., 2000. Evolutionary Robotics, the Biology, Intelligence and Technology of Self-Organizing Machines. MIT Press.
[37] Pfeifer, R., Scheier, C., 1997. Sensory-motor coordination: the metaphor and beyond. Robotics and Autonomous Systems 20 (2), 157–178.
[38] Pouget, A., Snyder, L., 2000. Computational approaches to sensorimotor transformations. Nature neuroscience 3, 1192–1198.
[39] Prince, C., Helder, N., Hollich, G., 2005. Ongoing emergence: A core concept in epigenetic robotics. In: Proceedings of the fifth international workshop on Epigenetic Robotics. pp. 63–70.
[40] Rochat, P., Striano, T., 1999. Emerging self-exploration by 2-month-old infants. Developmental Science 2 (2), 206–218.
[41] Rosander, K., von Hofsten, C., 2002. Development of gaze tracking of small and large objects. Experimental Brain Research 146 (2), 257–264.
[42] Roucoux, A., Culee, C., Roucoux, M., 1983. Development of fixation and pursuit eye movements in human infants,. Behaviour Brain Research 10, 133–139.
[43] Scassellati, B., 1999. A binocular, foveated active vision system. MIT, Artificial Intelligence Laboratory, Tech. Rep. A.I. Memo No. 1628.
URL `citeseer.ist.psu.edu/scassellati99binocular.html`
[44] Schaal, B., Hopkins, B., Johnson, S., 2005. Prenatal development of postnatal functions.
[45] Schlesinger, M., Amso, D., Johnson, S., 2007. Simulating infants' gaze patterns during the development of perceptual completion. In: Proceedings of the 7th International Conference on Epigenetic Robotics. pp. 157–164.
[46] Schwartz, E., 1977. Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. Biol Cybern 25 (4), 181—94.
[47] Scudder, C., Kaneko, C., Fuchs, A., 2002. The brainstem burst generator for saccadic eye movements. Experimental Brain Research 142 (4), 439–462, copy obtained.
[48] Sparks, D., 2002. The brainstem control of saccadic eye movements. Nature Reviews Neuroscience 3 (12), 952–964.
[49] Triesch, J., Teuscher, C., Deak, G., Carlson, E., 2006. Gaze following: why (not) learn it? Developmental Science 9 (2), 125–147.
[50] Tweed, D., Vilis, T., 1987. Implications of rotational kinematics for the oculomotor system in three dimensions. Journal of Neurophysiology 58 (4), 832–849.
[51] Zhang, Z., 2000. A flexible new technique for camera calibration. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22 (11), 1330–1334.