Modeling Human Bipedal Navigation in a Dynamic Three Dimensional Virtual Environment

Mark. D. Thomas[1,2], Daniel W. Carruth[1,2], Bryan Robbins[1,3],
John A. McGinley[1], and Alex Morais[1,3]

[1]Center for Advanced Vehicular Systems, [2]Department of Psychology,
and the [3]Department of Computer Science and Engineering
Mississippi State, Mississippi 39759
{mthomas, dwc2, bryanr, jam, amorais}@cavs.msstate.edu

Abstract. The current research sought to construct a computational model of human navigation for virtual three dimensional environments. The model was implemented within the ACT-R cognitive architecture [1]. The navigation model incorporates visual search, encoding object features and spatial relationships, motion, obstacle avoidance, and incidental visual memory.

Keywords: walking navigation, ACT-R, digital human model, incidental visual memory, visual search.

1 Introduction

One can likely stroll through a shopping mall and observe people walking while talking on cell phones, listening to music, or even chewing gum. Since people have the ability to do a wide variety of tasks while walking, people seem to be able to navigate without much thought. The apparent ease with which humans navigate begs one question for digital human modelers. *Why bother?* After all, end effectors can be used to motivate digital human models in virtual environments and many biomechanical and ergonomic constructs can be researched. However, a model of human navigation would be beneficial for other applications. For example, an architect may want to model an office building complete with digital human model (DHM) workers to test the emergency exits so changes could be made to the building design or escape route signs before construction begins. While such an application may seem unrealistic, researchers have claimed knowledge doubles every five to ten years in some scientific fields [2]. If digital human modeling advances at one-half the speculated rate, this and other diverse DHM applications are probably in the future.

1.1 Kinesthesis and Proprioception

Researchers have speculated navigation consists of motion and wayfinding with wayfinding defined as the cognitive aspects of navigation that do not involve motor actions [3], [4]. Proprioception is the sum of the kinesthetic and vestibular systems and it gives one the internal sensations of directional movement. Kinesthesis refers to the knowledge one has about ones body regarding the relationships between body parts due to sensations from muscles, tendons, and joints. The vestibular system relies on organs in the inner ear and provides feedback about one's orientation and movement. To navigate and return to roughly the same point of origin a blindfolded person could rely on proprioceptive cues by completing four series of paces and 90° turns.

In order to navigate from and return to a point of origin, one must have some strategy or internal representation of the point of origin. Sadeghiam, Kantardzic, Lozitskiy, and Sheta proposed several cognitive elements of wayfinding that parallel the psychological notion of spatial knowledge [4], [5], [6]. Researchers have also speculated about the existence of internal maps [7], [8], [9], [10], [11], [12]. While people may rely on internal representations, they may use external cues as well. For example, external cues can be used for navigation by following a sidewalk around the block to arrive at the initial point of origin, or a shopper may exit a store and depend on cues, such as a sign for a specific parking row, to locate his or her car. Rather than intentionally learning a landmark to be later recalled (e.g., a sign), the car could be located by recognizing a particular landmark, such as a streetlight with a bent pole, that was not intentionally learned when entering the store. The latter is an example of incidental visual memory.

1.2 Incidental Visual Memory

Incidental visual memory refers to memory for visual information that is not actively learned. In navigation, landmarks may be intentionally or incidentally learned in order to retrace a previously traveled route. Castelhano and Henderson investigated whether the intention to remember visual details impacted visual memory [13]. They had participants view scenes in an intentional learning condition and in an incidental learning condition (visual search). Participants remembered details of the scenes regardless of how the scenes were viewed initially. They proposed that memory for objects depended on whether the object was looked at, how often it was looked at, and the length of time it was viewed, not what instructions were given to the participant.

Using a conjunction visual search task, Williams, Henderson, and Zacks had participants count predefined target objects in a display of real-world distractor objects [14]. Williams et al. found that target objects were generally viewed more often and remembered better than distractor objects. Target memory rate was 85% while objects related to the target were remembered at approximately 60% and objects not related to the target were remembered slightly above chance. They also replicated Castelhano and Henderson's finding of a relationship between visual memory and viewing behavior for distractor objects [13].

The Williams et al. and the Castelhano and Henderson search tasks required participants to remain stationary and look at stimuli presented on computer monitors with the goal of searching a confined space [13], [14]. Other researchers have demonstrated incidental memory effects in real-world environments. Change blindness is a phenomenon in which people fail to detect changes in scenes if those changes occur during visual disruptions. Simons, Chabris, Schnur, and Levin had a confederate ask passersby for directions [15]. While receiving directions, a group of confederates passed between the direction giver and the person receiving the directions. While the direction received was occluded by the crowd, a member of the crowd exchanged a basketball with the direction receiver. Simons et al. reported that some participants were able to report visual details of the basketball although they did not report the exchange. Simons et al. found incidental memory for details of the exchanged object to be about 54%.

1.3 Visual Navigation

In order to navigate one must also have some knowledge of one's relationship to the objects in the environment and this knowledge has to be updated when moving through the environment [16]. Gibson speculated that the constructs of a *focus of expansion* and *optic flow* (or retinal flow) were the basis of human navigation [17], [18]. Focus of expansion refers to the region in the visual field at which an object appears to increase in size as one moves toward it. As one gets closer, the object covers more of the visual field. Using the focus of expansion to navigate can be accomplished by overlapping the focus of expansion on a desired goal and walking toward it [16], [17], [18], [19], [20]. (See Warren for a thorough review of focus of expansion and optic flow [19]).

Optic flow refers to the displacement of the environment on the retina as a person moves. Navigation can also be accomplished with optic flow by using a flow-equalization strategy in which the radial displacement of images on the retina is consistently maintained at proportional rates relative the visual field [16], [21]. In other words, distance from an object can be kept constant by maintaining a consistent movement of the object in the periphery of the retina. An example is walking parallel to a wall without looking directly at the wall.

Research has demonstrated that optic flow can be used for navigation and computer simulations have shown that optic flow is capable of providing the impetus for navigation [16], [22], [23], [24], [25]. Kim and Turvey proposed the *"units" of an optic array are projections of facets, faces, and surfaces of the environmental layout to the point of observation* [23]. Mathematically, optic flow is the relative velocity of points across the visual field as a person moves in the environment. Optic flow should operate whether one looks in the direction one is moving, or not, because optic flow depends on external environmental cues that change with the direction of movement. Using oscillating dots presented in linear and rotating fashions, Regan and Beverley demonstrated differences between processing linear and rotating stimuli, which suggested unique processing mechanisms may exist for the curl of velocity (i.e., vorticity) and the divergence of velocity (i.e., dilation), which in turn implies special processing mechanisms may exist for optic flow [26]. Regan and Beverley further speculated the mechanism for processing optic flow may include an ability outside one's awareness that parallels vector calculus to determine and track the extent of optic flow changes when moving.

Other research has brought the importance of optic flow into question. Cutting, Reading, and Wang examined whether people veered in the direction they looked

[27]. Cutting et al. had participants walk straight while looking to the side in an illuminated condition and in a darkened condition. They found that people veered in the direction looked. Cutting et al.'s results appear to contradict optic flow theory. Because participants never had the opportunity to look at a destination goal, an argument could be made the veering behavior resulted from participants using an egocentric-direction strategy for navigation [16], [28], [29]. During egocentric-direction strategies, a person is believed to visually mark, or tag, a goal and walk toward it. Cutting et al. found no difference between the extent of veering and the lighting conditions. If visual-based navigation depended on optic flow, one would have anticipated differences because almost no visual input was available in the darkness condition. Conversely, if navigation depended on checking headings against a focus of expansion, one would have expected the veering behavior to be the same. However, navigation in low illumination environments and environments with few external cues has been demonstrated and sighted individuals may rely on more than vision to navigate [29], [30], [31]. Participants in Cutting et al.'s research may have used navigation strategies that depended very little, or not at all, on vision (e.g. kinesthesis and proprioception) [27].

2 Incidental Visual Memory and Navigation

The following data were previously reported by Thomas et al. [32]. Space constraints do not allow a complete recapitulation. This brief summary is included to provide some details on the methodology and results of the human data that the current modeling effort is attempting to match. Please consult the original study for a better understanding of the methodology.

2.1 Participants

The participants were 12 males with a mean age of 19.08 years (R = 18 - 24) who received monetary compensation for their participation and were in self-reported "average" or better health status. No participants reported a history of head or brain anomalies (traumatic injuries or illnesses). Participant visual acuity and color vision was screened. Immediately prior to their participation, the participants took part in a motion capture study that required wearing a portable motion capture suit and the eye tracker.

2.2 Materials, Apparatus, and Procedure

Materials. The navigation task included three rooms: the laboratory, a workshop, and a break area. See Figure 1 for an illustration of the layout. The laboratory served as the starting and ending point of the navigation task. The workshop floor dimensions were 230 feet × 49.5 feet with a ceiling height of 35 feet in one area and a dropped ceiling with a height of 9.5 feet in an area 7.5 feet from the wall with various laboratory entry doors spanning the entire length of the workshop. A 30-feet section spanning the breadth of the workshop at the area farthest from the lab also had the dropped ceiling. The break area consisted of an open floor plan with a glass block curved wall.

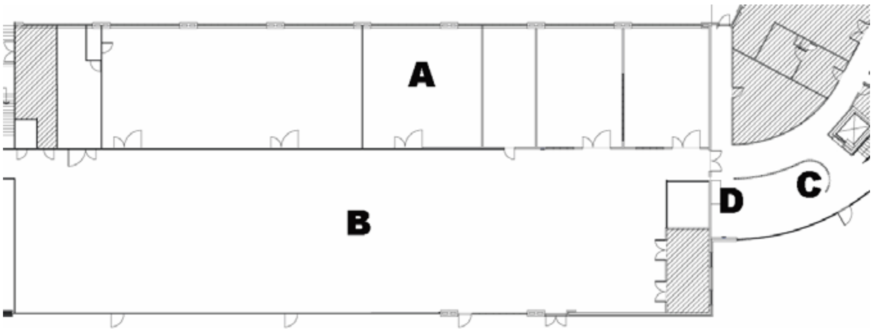

Fig. 1. Floor Layout: A. Human Factors and Ergonomics Lab, B. Workshop, C. Curved glass wall, D. Soft drink vending machine

The memory test stimuli were composed by taking advantage of natural occlusions in the workshop. The photographs were constructed so that none of the manipulated objects overlapped in the photographs. When possible, the photographs were taken from 3 different angles of incidence (left, right, and head-on). The images were reduced and cropped to 300 x 300 pixels. Three versions were photographed of each viewing angle when possible: (a) the original, unchanged scene, (b) with an object added to the scene, (c) and with an object removed from the scene. A fourth version of each scene was created by digitally changing the color of an object(s) in the scene. Thus, each scene had as many as 3 views and each view had as many as 3 foils which is a possible 9 foils for each scene. However, in order to prevent overlap of the manipulated scene regions and due to naturally occurring obstructions, composing 3 views of each scene was not always feasible, nor was adding or removing objects always feasible.

Apparatus. Right eye point-of-gaze was recorded with an Applied Science Laboratories Mobile-Eye tetherless infrared-video-based eye tracker. The eye tracker camera output at 60 Hz and the video capture was 29.97 frames per second. Therefore, the actual sampling rate is approximately 30 Hz. Memory test stimuli were displayed at a resolution of 800 × 600 pixels × 24-bit color on a flat screen LCD computer monitor viewed at approximately 75 cm (35.43° × 26.57° of visual angle). Test scenes were 300 × 300 pixels and subtended approximately 8.89° × 8.89° of visual angle at 75 cm viewing distance.

Procedure. Participants were escorted to the laboratory without passing through the portion of the workshop they would later be asked to navigate. Once in the laboratory, informed consent was obtained. Demographic data, anthropometric data, and a brief medical history were collected. Visual acuity and color vision were screened. Participants were fitted with a portable motion capture suit and the eye tracker. We interacted with participants for approximately 1 hour during another experiment before the navigation task began. Participants wore the eye tracker for the other tasks as well as the navigation task. Participants were not told about the memory test until immediately before test administration and they were not told we were investigating navigation until debriefing. We gave each participant change to buy a soft drink from a vending machine and told the participant to buy a drink of his choice. We pointed in the general direction of the vending machine and instructed the participant to, "Just go that way until you find the drink machine.” If participants sought clarification, the researcher repeated the hand gesture and verbal instructions. No other clarification was offered.

Navigation consisted of walking out of the lab through a door and turning left, then walking 110 feet through a closed pair of double doors. Although the break area was not visible through the doors, it was visible through a small opening in the wall immediately after exiting the workshop via the double doors. A barricade was placed to prevent participants from using the immediate path. Thus, participants had to walk an additional 30 feet around the curving glass wall, turn 90° and walk another 25 feet to reach the vending machine. After interacting with the vending machine, participants had to return around the glass block wall, pass through the double doors, walk partially through the workshop, and pass through the lab door.

Upon returning to the lab, participants were immediately given an unannounced two-alternative forced-choice recognition test. Participants were tested on 18 scenes from the workshop. Test objects were presented to the right and left of a fixation point. The foil item was the same scene with either an item added to the scene, removed from the scene, or with a color replaced for an object in the scene. No scene was tested more than once for any participant. Task instructions were presented on the initial memory test screen and responses were input by participants with a standard button box. Participants were instructed to guess if they were uncertain. No feedback was given.

2.3 Analyses and Results

Video coding. All videos were independently coded by a senior undergraduate student and the primary author of the current paper. From a possible 31,944 coded frames, 26,325 were agreed upon. Interobserver agreement was 82.4% and the disagreements were resolved by conference. The reported eye movement and viewing behavior results are for 9 participants only. Two were excluded based on frame drop rates ($\geq$ 90%) and the third file was corrupted. Percentages of the lengths of time spent looking at a focus of expansion were calculated by adding all the instances of looking at what appeared to be a focus of expansion for the entire navigation sequence minus the time spent interacting with the vending machine.

Scene viewing behavior was erratic. Participants often appeared to make ballistic saccades from one object to another within a scene. If this was the case, apparent fixations consisted of a single frame, but without at least 2 frames, an apparent fixation may have been a sample taken in mid-saccade. All calculations are based on a minimum of 2 consecutive frames and are believed to underestimate actual viewing behavior. Of the 9 videos retained for analyses, the average rate of frame drop was 22.17% and, other than eye blinks, it appears most of those were due to participants looking beyond the angle capabilities of the camera. Despite possible frame rate concerns, some objects were viewed as many as 53 frames (nearly 2 s).

Incidental Visual Memory. The average time spent navigating through the environment that was tested for memory was 60.5 s. (R = 44.47 - 78.88 s). The average time spent viewing the tested scenes was 6.9% (calculated by using a minimum of 2 consecutive frames). Memory test results indicated the incidental memory for the scenes was 58.82% [$t(11) = 2.39$, $p = .036$, $SE = .037$] with a standard deviation of 12.79%. Performance for 1 participant was more than 2 standard deviations below the mean and performance for another was more than 2 standard deviations above the mean. Both were included in the reported memory results. Performance excluding the two possible outliers ranged from 47.06% to 64.71%.

Navigation Behavior. The average total time spent navigating (total time from start to finish excluding the machine interaction time) was 118.4 seconds (R = 86.2 - 182.7 s). The 182.7 s time is inflated because the gentleman walked past the break area and into other areas of the building. Although 2 participants did not view any of the same scenes during the return trip through the workshop, an average of 2.55 objects (4.7%) viewed during the initial trip through the workshop were viewed again during the return trip through the workshop. Six participants viewed an initial object immediately outside the lab, but only 2 viewed the object again when returning to the lab.

Vertical architectural elements were viewed more often than any other aspect of the environment with 11.24% of the total navigation time spent viewing walls, columns, and other vertical spatial boundaries. Participants appear to have defined an average of 6 areas of reference that may have been foci of expansion. Including the initial reference gaze, participants averaged 21.1 looks at prospective foci of expansion for an average of 9.9% of the time spent while walking. In some instances the possible foci of expansion included walls and these instances were counted as both possible foci of expansion and vertical spatial boundary looks. The subjective ranges of times per looks at a possible focus of expansion ranged from a short 0.07 s glance to a long 5.2 s stare. The average reference look at a possible focus of expansion was 0.52 s.

3 Navigation Model

The navigation model is a computational model implemented within the ACT-R cognitive architecture [1]. The ACT-R cognitive architecture is a production system symbolic architecture that incorporates models of memory, learning, perception, and action. Models of human performance constructed within ACT-R specify if-then rules that coordinate the goal, memory, perception, and action modules. The ACT-R architecture and incorporated modules are formal implementations of psychological theories.

3.1 The Tasks

Navigation. The navigation task can be divided into direction inference, foci of expansion target identifications and reference looks, obstacle avoidance, and incidental visual memory. When using a focus of expansion, the actor selects a heading target to serve as a focus of expansion and moves toward that target. The actor periodically checks the focus of expansion to verify the heading and once the navigation target is reached, another navigation target is selected.

Obstacle avoidance. When an obstacle blocks the participant’s desired path, the space at the extents of the obstacle can be tested to determine if a viable path exists to attempt to bypass the obstacle. In the navigation task, the only open space is at the left-most extent of the glass wall. The participant navigates to this space with a goal of finding an opening in the wall. In this case, the participant does find an opening. In a more complex, maze-like environment, other spatial based memory strategies would be needed.

Incidental Visual Memory. As the actor navigates through the environment and searches for the goal target (e.g., vending machine or lab), the actor is not only looking at and encoding obviously relevant objects but also other interesting objects in the environment. The shifting of attention to “irrelevant” objects is based on a combination of bottom-up and top-down factors. The visual features of some objects are particularly salient and will attract attention. Also, the visual features driving the participant’s search for the goal (i.e. color: red, category: box) will influence what objects are considered worthy of focused attention. The current model simulates these factors via feature-based search of the visual array.

3.1 The Sub-Tasks

Visual Search. Search is modeled as a feature-based conjunction search of the visual array [33]. As suggested by Williams, et al., the actor has some concept of a goal target (e.g., a soft drink vending machine) [14]. The actor searches the visual array for the features that are associated with their mental representation of the goal target. In this task, the actor has not seen the actual goal target and is assumed to be using some exemplar representation.

Encoding. The encoding of visual objects is accomplished by integrating features into chunks. The subsequent chunks are stored in declarative memory. Once visual features are encoded as objects, the model can compare them to representations of the goal target or evaluate them as potential targets for navigation.

Spatial Relationships. In navigation, the spatial relationship between the self and the environment is particularly important [16]. Spatial information is encoded from the visual array as part of the visual object chunk. This spatial information is egocentric in nature. The visual object chunk encodes the observer's bearing to and distance from the encoded object. An allocentric representation of the environment is built by constructing object-to-object relationships from the egocentric spatial relationships.

4 Conclusion

The current research seeks to construct computational models of human navigation in virtual three dimensional environments. The current model, within the ACT-R cognitive architecture [1], incorporates visual search, encoding of object features and spatial relationships, motion, obstacle avoidance, and incidental visual memory.

The navigation model is a computational model implemented within the ACT-R cognitive architecture which is a production system symbolic architecture that incorporates models of memory, learning, perception, and action [1]. The visual search implementation is a feature-based conjunction search of the visual array. The encoding of visual objects is accomplished by integrating features into chunks that are stored in declarative memory. Spatial information is encoded from the visual array as part of the visual object chunk. The spatial information is egocentric and the visual object chunk encodes the observer's bearing to and distance from the encoded object. The allocentric representation of the environment is constructed from the egocentric object-to-object spatial relationships. Navigation tasks used multiple foci of expansion target identifications and reference looks, obstacle avoidance, and incidental visual memory. The extents of the object were used to select a new navigation targets to bypass obstacles. Information about interesting objects in the environment is encoded.

References

1. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An Integrated Theory of the Mind. Psychological Review, 111 (2004). 143-166.
33. Anderson, J.R., Matessa, M, Lebiere, C.: The Visual Interface. The Atomic Components of Thought. In J.R. Anderson and C. Lebiere eds. Erlbaum. Mahwah, NJ. (1998). 143-166.
13. Castelhano, M.S., Henderson, J.M. Incidental Visual Memory for Objects in Scenes. Visual Cognition, 12 (2005). 1017-1040.
27. Cutting, J.E., Reading, W.O., Wang, R.F.: Walking, Looking to the Side, and Taking Curved Paths. Perception & Psychophysics, 64 (2002). 415-425.
3. Darken, R., Sibert, J.: Navigating Large Virtual Spaces. International Journal of Human-Computer Interaction, 8 (1996). 48-72.
21. Duchon, A.P., Warren, W.H.: A Visual Equalization Strategy for Locomotor Control: Of Honey Bees, Robots, and Humans. Psychological Science, 13 (2002). 272-278.
17. Gibson, J.J.: Perception of the Visual World. Houghton-Mifflin, Boston, MA (1950).
18. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton-Mifflin, Boston, MA (1979).
28. Harris, J.M., Bonas, W. Optic Flow and Scene Structure do not Always Contribute to the Control of Human Walking. Vision Research, 42 (2002). 1619-1626.
22. Harris, M.G., Carre, G.: Is Optic Flow Used to Guide Walking While Wearing a Displacement Prism? Perception, 30 (2001). 811-818.

7. Hirtle, S.C., Jonides, J.: Evidence of Hierarchies in Cognitive Maps. Memory & Cognition, 13 (1985). 208-217.
30. Hollands, M.A., Patala, A.E., Vickers, J.N.: "Look where you're going!": Gaze Behavior Associated with Maintaining and Changing the Direction of Locomotion. Experimental Brain Research, 143 (2002). 221-230.
8. Huttenlocher, J., Hedges, L.V., Duncan, S.: Categories and Particulars: Prototype Effects in Estimating Spatial Location. Psychological Review, 98 (1991). 352-376.
23. Kim, N.-G., Turvey, M.T.: Eye Movements and a Rule for Perceiving Direction of Heading. Ecological Psychology, 11 (1999). 233-248.
31. Land, M.F., Tatler, B.W.: Steering With the Head: The Visual Strategy of a Racing Driver. Current Biology, 11 (2001). 1214-1220.
9. McNamara, T.P.: Mental Representations of Spatial Relations. Cognitive Psychology, 18 (1986). 87-121.
2. Mitchell, M., Jolley, J.: Research Design Explained. 4th edn. Harcourt College Publishers, Fort Worth, Texas (2001).
26. Regan, D., Beverley, K.I..: Visual Responses to Vorticity and the Neural Analysis of Optic Flow. Journal of the Optical Society of America, 2 (1985). 280-283.
11. Rosenbaum, R.S., Ziegler, M., Winocur, G., Grady, C.L., Moscovitch, M.: "I have often walked down this street before": fMRI Studies on the Hippocampus and Other Structures During Mental Navigation of an Old Environment. Hippocampus, 14 (2004). 826-835.
29. Rushton, S.K., Harris, J.M., Llyod, M.R., Wann, J.P.: Guidance of Locomotion on Foot Uses Perceived Target Location Rather Than Optic Flow. Current Biology, 8 (1998). 1191-1194.
4. Sadeghiam, P., Kantardzic, M., Lozitskiy, O., Sheta, W.: The Frequent Wayfinding-Sequence (FWS) Methodology: Finding Preferred Routes in Complex Environments. International Journal of Human-Computer Studies, 64 (2006). 356-374.
15. Simons, D.J., Chabris, C.F., Schnur, T., Levin, D. T.: Evidence for Preserved Representations in Change Blindness. Consciousness and Cognition, 11 (2002). 78-97.
10. Stevens, A., Coupe, P.: Distortions in Judged Spatial Relations. Cognitive Psychology, 10 (1978). 422-437.
12. Taylor, H.A., Tversky, B.: Descriptions and Depictions of Environments. Memory & Cognition, 20 (1992). 483-496.
32. Thomas, M.D., Carruth, D.W., McGinley, J.A., Follett, F. Task Irrelevant Scene Perception and Memory During Human Bipedal Navigation in a Genuine Environment. In Proceedings of the 25th Army Science Conference, Orlando, FL (2006).
5. Thorndyke, P.W., Hayes-Roth, B.: Differences in Spatial Knowledge Acquired From Maps and Navigation. Cognitive Psychology, 14 (1982). 560-589.
6. Thorndyke, P.W., Stasz, C.: Individual Differences in Procedures for Knowledge Acquisition From Maps. Cognitive Psychology, 12 (1980). 137-175.
16. Turano, K.A., Lei Hao, D.Y., Hicks, J.C. Optic-flow and egocentric-direction strategies in walking: Central vs peripheral visual field. Vision Research, 45 (2005), 3117-3132.
19. Warren, W.H. Jr.: Visually Controlled Locomotion: 40 Years Later. Ecological Psychology, 10 (1998). 177-219.
20. Warren, W.H., Kaye, B.K., Zosh, W.D., Duchon, A.P., Sahuc, S.: Optic Flow is Used to Control Human Walking. Nature Neuroscience, 4 (2001). 213-216.
24. Wilkie, R.M., Wann, J.P.: Driving as Night Falls: The Contribution of Retinal Flow and Visual Direction to the Control of Steering. Current Biology, 12 (2002). 2014-2017.
25. Wilkie, R.M., Wann, J.P.: Eye-Movements Aid the Control of Locomotion. Journal of Vision, 3 (2003). 677-684.
14. Williams, C.C., Henderson, J.M., Zacks, R.T. Incidental Visual Memory for Targets and Distractors in Visual Search. Perception & Psychophysics, 67 (2005), 816-827.