Visuovestibular-Based Gaze Control Experimental Case

José Prado, Jorge Dias
jaugusto@isr.uc.pt, jorge@isr.uc.pt

***Abstract*— Acquiring data from human behavior when inserted into a given environment and using this information as prior knowledgement into our artificial system was our purpose. Of course human beings are very complex and might react to almost any kind of stimulus based on different semsations and prior knowledgement what turn us impossible to be fully predicted or imitated by a computer algorithm. In other hand if we restrict the human environment conditions, we also can restrict the scope of possible human reactions. Human beings have natural sensors to capture sempsations of it's five senses. Moreover human have also the vestibular system that is inside the internal ear and give accelaration signals to be interpreted by the brain. For sure humans have much more sensors than these ones mentioned before and we could be describing human body but the point aimed here is that finally we will certainlly want to enshort the scope of interest in the behalf of human senses. The first restriction here is that we want to imitate reactions based on sight(vision) and vestibular system only. To do so, we must create an environment where the other human senses have as low influence as possible in the experiments. Inside vision we can say that we don't care about color, we will not do object segmentation, neighter object or face recognition.**

I. Introduction

A. Bayesian Model of Visuovestibular-Based Gaze Control

The purpose is about reaching to a feasible Bayesian model for a robotic gaze control following the ideas presented on [1] and [2]. From the camera image we extract Fd^t and Fa^t which are the direction and the amplitude of the mean flow. From the IMU we get the angles (Roll, Pitch and Yaw) that are further shown in this paper as R^t (for Roll), P^t (for Pitch), and Y^t (for Yaw).The actuator control acts based on current angular position and on instantaneous flow information of the system. The pan-tilt unit are controlled with combined commands for target position and velocity. The motors move to the desired target position with the selected velocity and stop. The motor model takes this into account by having the current motor command depending on the current state and also on the probabilistic table filled out with the human reaction information.

The following variables were be used:

- S_t: is a tuple with the following four variables transformed in possible motor reactions
  - R^t: (roll) angle of the human-reaction for a given state
  - Y^t: (yaw) angle of the human-reaction for a given state
  - Fd^t: direction of the vector of the mean flow (comes from vectored product between u and v) (Radians)
  - Fa^t: amplitude of the vector of the mean flow
- M_t: is a movement variable with the following scope (UP, DOWN, LEFT, RIGHT, STOP)
  - The five states of M_t are concluded by doing atomization of the raw values in the following variables
    - ∗ pan motor velocity: $\mathcal{P}_\omega$ — pan motor target position: $\mathcal{P}_\theta$
    - ∗ tilt motor velocity: $\mathcal{T}_\omega$— tilt motor target position: $\mathcal{T}_\theta$
- H_t: is the human reaction to be learned (UP, DOWN, LEFT, RIGHT, STOP)

To simplify notation, state variables are grouped in a vector $S = (R^{0:t}, P^{0:t}, Y^{0:t}, Fd^{0:t}, Fa^{0:t})$ and motor variables are considering to be in the range U,D,L,R,S after atomization from $M = (\mathcal{P}_\omega, \mathcal{P}_\theta, \mathcal{T}_\omega, \mathcal{T}_\theta)$. The bayesian program that show the relation between these variables is shown in fig. 1.

Fig. 1. BayesianProgram

Program
- Description
  - Specification
    - Variables:
      - $S^{0:t} = (R^{0:t}, Y^{0:t}, Fd^{0:t}, Fa^{0:t})$: state variable
      - $M^{0:t}$: motor movement variables with scope {U,D,L,R,S}
      - $H^{0:t}$: human reaction variables with scope {U,D,L,R,S}
      - $Fd^{0:t}$: flow direction (Rad)
      - $Fa^{0:t}$: flow magnitude (pixels)
    - Decomposition:
      - $P(S^t \wedge M^t \wedge H^t \wedge Fd^t \wedge Fa^t =$
      - $P(R^t.P(y)^t)$
      - $.P(H|R \wedge y)$
      - $.P(Fd|H \wedge R \wedge y)$
      - $.P(Fa|Fd \wedge H \wedge R \wedge y)$
      - $.P(M|Fa \wedge Fd \wedge H \wedge R \wedge y)$
    - Parametric forms:
      - $P(M^t) = P(H^{0:t}|S^t)$
  - Identification:
    - Bayesian human-learned gaze control.
- Question:
  - $P(M^t|S^t \wedge H^{0:t})$

II. Experimental Paradigm and protocols

A. Apparatus and Stimuli

We want to use a head mounted device (HMD) to pass the visual stimulus from the robot to the human subject. However currently we used the input from the keyboard reflexes when the subject looks to the screen. Our robotic head (fig.2) is a common platform consisted by many sensors, but basically those that we are using are a stereo camera, a pan-tilt unit and a inertial sensor, those are attached statically and being so every motor command sent to the pan-tilt unit will reflect on IMU motion and also in the camera images, consequently in the calculated optical flow and inertial data.

Fig. 2. Robotic Head

For software, each device has a correspondent module and they were integrated with our robotic software platform derived from [3].

B. Subjects

Five human subjects with normal working visual and vestibular systems. In those subjects with visual distortion this should be compensated by using glasses or lens, thus the distortion perform no impact to the experiment. We will mix male and female according to avaiability of the persons. The subjects should be at least three NAIF and two authors.

C. Protocol

By using the HMD to give the robot images to the human eyes the visual conection becomes direct between the robot and the human. We can not inject artifitial inertial sensor data into human brain. Thus, what is possible to be done is a indirect correlation where during the tests, the human will use it's own vestibular system while the robot will use the artificial inertial system. Gray scale images are the input, with several visual detectable features on the environment. Visual Features are necessary by the human brain to have notion of motion. If a human is moving sidelly in front of a perfectlly white wall, once the accelaration stabilizes subject will have no sensation of motion. However if this same wall is full of visual detectable features, human will naturally detect the motion only by the visual influence. We have the same response on artifitial optical flow algorithms, and that's why we are interested on considering the optical mean flow as a artificial visual ego-motion notion measurement variable.

III. Results and Conclusion

A first version of human-learning was implemented, using keyboard to control the robot while monitored by a human (human in the loop way of learning like in [4]). We still want to improve our way of learning by buiting a helmet equiped with camera and IMU and then detecting real human neck movements.

Consider that we numbered our random variables as follows:

1) is $Roll^t$, subvariable of Imu^t variable
2) is $Pitch^t$, subvariable of Imu^t variable
3) is Yaw^t, subvariable of Imu^t variable
4) is Fd^t, the Flow Direction
5) is Fa^t, the Flow Amplitude

The learned table is a 4D probability table with dimentions [36x5x10x5] in our test, we plotted this in five 3D graphics in order to be possible to visualise them.

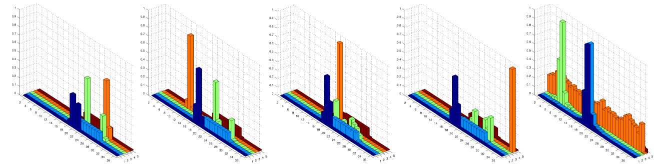

Fig. 3. Probability Table - Learned data for UP, DOWN, LEFT, RIGHT and STOP movement

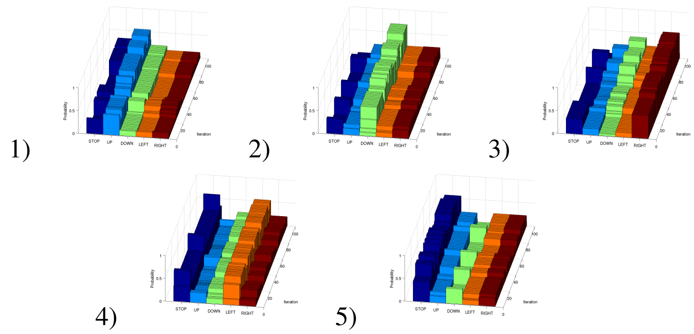

Fig. 4. System reaction - 1) Falling down in free fall "forever" (simulation) 2) Launched up as a rocket to the sky "forever" (simulation) 3) Transladating to left side "forever" (simulation) 4) Transladation to right side "forever" (simulation) 5) Stoped (Imu^t vary even stoped because of Mag^t mainlly)

It is possible to observe that for the UP, DOWN, LEFT and RIGHT movements, the main cathegorizing random variable is Fd^t, in the other hand for the STOP movement Fd^t is very confusing, thus Imu^twill be much more usefull cathegorizing this decision.

a) Testing the reaction of the system: Fake stimulus were injected into the system to measure if the robot's reaction would be like expected for that stimulus. As human trained the system, we know (aproximadelly) which stimulus to create and which reaction to expect. For example if we train a walking robot not to fall from a step, we can put a step in front of it and our expectation will be that the robot do not fall. In our case we trained the head to be centered and then we give stimulus simulating that the head would moving to one or another side "forever" during each test. We also gave stimulus for the system to believe the head was flighting up like a rocket and also falling down in free fall. It was performed 100 tryals with different stimulus for each espected reaction.

One stimulus for each reaction would be the trivial case to cathegorize, but for this preliminary results we had aproximatedlly 98% of correct decisions in 500 different stimulus to be cathegorized in 5 movements.

References

[1] Pierre Bessiere, Juan-Manuel Ahuactzin, Kamel Mekhnacha, and Emmanuel Mazer. *Bayesian Programming Book*. 2006.
[2] Pierre Bessiere, Christian Laugier, and Roland Siegwart. *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*. Springer, 2008.
[3] Michael Montemerlo, Nicholas Roy, and Sebastian Thrun. Perspectives on standarization in mobile robot programming: The carnegie mellon navigation (carmen) toolkit. 2003.
[4] Xavier Perrin, Ricardo Chavarriaga, Roland Siegwart, and Jose del R. Millan. Bayesian controller for a novel semi-autonomous navigation concept. *ECMR*, 2007.