Measurement of Human Concentration with Multiple Cameras

Kazuhiko Sumi, Koichi Tanaka, and Takashi Matsuyama

Graduate School of Informatics, Kyoto University, Kyoto 606–8501, Japan,
sumi@vision.kueee.kyoto-u.ac.jp,
WWW home page: http://vision.kuee.kyoto-u.ac.jp/

Abstract. We propose a new method to estimate human change of concentration from multiple camera views of the human. In our method, human state of concentration is observed as self-load, defined as energy injected in a period to keep and manipulate his/her body. If a person is concentrating to a certain task, he/she will brace himself/herself for better results, and energy consumption will increase. To confirm our idea, we developed a method to calculate self-load from multiple view of the human. We conducted an experiment in which test subjects have different level of complexity of task. Self-load of the subjects showed the positive correlation with the complexity of the task. We have convinced that self-load can be used to characterize the concentration of person being observed.

1 Introduction

One of the big difference between human-to-human interaction and human-to-computer interaction is timing. We human can measure timing to start conversation with another human. Measuring timing is a very sophisticated social action, which is not achieved by a computer yet. To realize such a behavior, it is important to have a good sensing systems to observe the human action and his/her internal state. In this research we focus on sensing human internal state, such as concentration, interest and frustration.

So far, many studies on human observation were carried out. Most of them are recognizing intentional signal in communication. Such researches include gesture recognition, facial expression recognition, lip reading, and voice recognition. Those methods are quite reasonable when the person is already interacting with a computer. On the other hand, if a person is not involved in communication, but is engaging in other personal jobs, the channel of communication is not established. In such cases, a computer should estimate the human interest, intention and feeling through one-way observation. This will often happen when assisting a human involved in a work, such as driving a car, operating a machine, traveling in an unfamiliar places, looking for something, and so on. If the person is doing something in concentration, it is not a good manner to offer help. But if the person is wondering, it might better to support him/her. Thus observing a human without interaction is a challenging subject.

Human internal state such as stress or frustration can be measured through a various sensors. Picard developed wearable sensors for human observation[4] using galvanic skin response, blood volume pressure, and electromyograph. Fernandez applied the system to measure frustration of a human using a computer[5]. However, wearing such a device may not be comfortable, and sensing from outside of a human is preferred. Mota analyzed learner's interest level by measuring 2D pressure distribution between a human and a chair under the human[6]. This system provides better comfort, but, still need to be contacting toward something. To realize non-contact human observation in a free space, we are interested in observing a human from cameras.

Cameras can be used to observe a human through various cues of human behaviors, such as, eye and sight[8], expression[7], gesture[2], and body posture[3]. To understand a human better, it is preferred to get a close up view of the human. However, in a ordinary room, it is not easy to get a close up view of his/her face all the time. Thus we are interested in observing a human only with cameras with wide field of view.

In this research, we estimate the concentration, in other words, how much is the human is occupied by the task he is doing. We use multiple cameras surrounding the human to be observed. To achieve a quantitative analysis we introduce new measure, which is referred to as self-load. Self-load is a energy consumption per unit time. The energy consumption is derived from static and dynamic posture of the observed human.

In the following section, first we define self-load and its derivation from his/her posture in Section 2. Then, we explain the pose estimation method using multiple cameras in Section 3. To confirm the effectiveness of our proposal, two series of experiments are carried out in Section4. Finally, we conclude our proposal in Section 5.

2 Calculation of Self-load using Cylindrical Human Body Model

In human body condition, compared with the state of doing nothing, thus no stress in his/her muscle, the work load which supplies the energy per unit time to the present body expression is calculable. The amount of energy consumption is referred to as self-loads. Although self-load is a physical amount of energy and is not a mental state, it is reasonable to assume that there is a high correlation between physical energy consumption and mental concentration.

First, the amount of energies needed about maintenance of a certain body posture and operation is defined by potential energy, movement energy, and posture maintenance energy. Posture maintenance energy is the energy required in order to maintain the posture and to keep muscles tension. If the joint is in the neutral position, posture maintenance energy is zero. It is expressed with the difference from the neutral state. If a person is putting his/her part on a structure, posture maintenance energy should be deducted, because he/she can save stress of his/her muscle. However, it is difficult to measure a force between

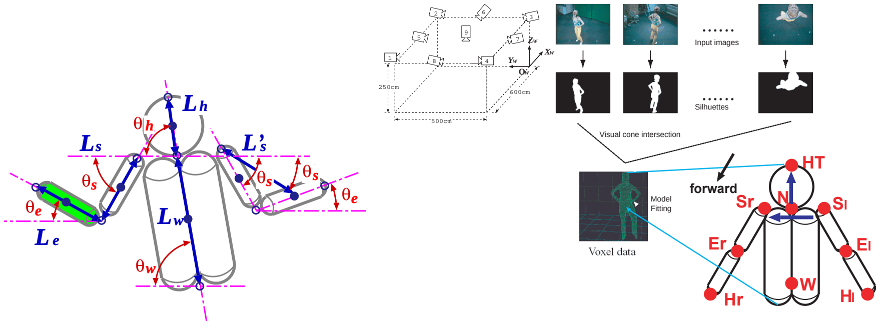


Fig. 1. A cylindrical human body model and parameters of its parts (left), and schematic diagram of human upper pose estimation from multiple camera images via 3D volume of a human (right)

the body and the structure, we will not consider the case in this research. Instead, we only treat a human with stable supporting energy.

In this chapter, we will apply Hill's muscle model[9], which considers an actuator pulling two springs with dumper, to calculate the sum of the energy consumed by a body part at a certain angle. Also, we assume that muscle actuation is iso-tensional, and the tension is only caused by gravity. This assumption is not satisfied when a person is pressing or pulling a hard structure with his muscle or when he is stressing his body. We don't consider such invisible cases in this study. Thus, we will consider visible part of posture maintenance energy which is expressed by the gravitation moment on a body part.

We model a human body by a set of cylinders shown in Figure. 1(left). A unit part of the body model is a single cylindrical part. Instantaneous self-load of the unit sl_e, lower arm for example, is the sum of potential energy sl_e^p, motion energy sl_e^m, and posture maintenance energy sl_e^k shown in Equation 1.

$$sl_e = sl_e^p + sl_e^m + sl_e^k = m_e g h_e + \frac{1}{4} m_e L_e \omega_e(t)^2 + sl_e^k \tag{1}$$

where, g, L_e, m_e, h_e, $\omega_e(t)$ are G-forces, length of the arm, mass of the arm, height of the arm center from the lowest position, and angular velocity of the arm at time t, respectively. Posture maintenance energy of the lower arm sl_e^k is:

$$sl_e^k = m_e g \frac{L_e}{2} \cos(\frac{\theta_e}{2}) - sl_{e\ init}^k \tag{2}$$

And its initial value is $sl_{e\ init}^k = \frac{1}{2} m_e g L_e$. We can calculate self-load of upper arms sl_{sl}, sl_{sr}, and that of chest sl_w in the same way. Instantaneous self-load of the whole upper body sl is the sum of self-load of all the parts:

$$sl = sl_{el} + sl_{er} + sl_{sl} + sl_{sr} + sl_w \tag{3}$$

Self-load of the body SL is defined by the temporal average of the instantaneous self-load during the time interval T:

$$SL = \frac{1}{T}\sum_{t=0}^{T} sl \qquad (4)$$

3 Pose Estimation from 3D Volumetric Representation

We estimate the pose parameters required for self-load calculation from the 3D volumetric representation. The 3D volumetric representation is derived by the intersection of visual cones[11]. For each visual cone, its top vertex is the camera center and its base plane is the silhouette of the body taken by the camera. This step is show in Figure.1(right). The model of upper body consists of 6 parts.

In the fitting process, the characteristic point of the body, head peak $\mathbf{HT}$ is searched first. Then from $\mathbf{HT}$, center of neck $\mathbf{N}$, right shoulder $\mathbf{S_r}$, and left shoulder $\mathbf{S_l}$ are searched in this order. From the neck $\mathbf{N}$, trunk of the body is scanned and waist center $\mathbf{W}$is located. From the shoulders $\mathbf{S_r}$ and $\mathbf{S_l}$, each arm is scanned toward hand, and elbow joint $\mathbf{E_l}$, $\mathbf{E_r}$, and hand tips $\mathbf{H_l}$, $\mathbf{H_r}$ are located.

4 Experiment

To confirm our idea, we conducted three experiments. First, we have examined the pose estimation accuracy of our method using life size figure of a human. Our experiments ware carried out in a laboratory shown in Figure. 1(left) with 9 cameras. This multiple camera system was calibrated in advance, and its voxel resolution is 2cm, its frame rate is 9 fps. The pose estination result shows the worst angular error of arms, waist, and head are 13.8deg, 7.0deg, and 6.0deg respectively. This error is acceptable for self-load estimation. However, the fitting algorithm sometimes failed when arms are contacted in parallel to trunk of the body. More robust fitting algorithm are required for future work.

4.1 Self-load measurement

In this experiments, we evaluated correlation between task complexity and self-load measure. We designed two scenarios. Each of them has three different levels of complexity. Total 7 subjects participated in the experimnets. The set up is shown in Figure 2(left). Due to limitation of the 3D reconstruction space, the pose of the subject is limited to a sitting pose and we couldn't fit lower body in this setup.

Before the real experiment, the subject are required to wait until the experiment is ready. During this period, which is 28sec long, a TV program is displayed on the PC screen and neutral pose are measured. From the measurement, we calculated the initial self-load for each subject.

In the first experiment, which we refer to as "clap test", a white box in the black background on another PC screen appeared 4 times during the 28sec

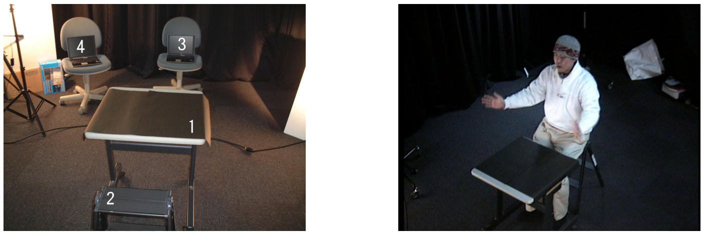


Fig. 2. Self-load measurement set up. (left) – 1: Desk, 2: Stool, 3: Task Screen, 4: TV Screen –, and a snap shot of the clap test

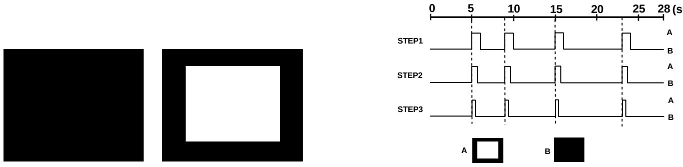


Fig. 3. Screen shown to the subject during clap test

session. A subject was requested to clap his/her hands, when he/she recognize a white box before it disappears. The duration of white box is same in a session but it becomes shorter, shown in Figure. 3, and the task becomes harder. Each subject performs three sessions and self-load is measured for each session. Figure. 2(right) shows one of the snapshot during this experiment.

In the second experiment, which we refer to as "catch test", a button with "START" mark and a moving box with "CATCH" mark, shown in Figure. 4(left), are displayed on the screen. A session starts when the subject clicks the mouse on "START", then "CATCH" starts running on the screen. The subject is requested to follow the moving box and click the mouse when it is on "CATCH" mark. The subject is requested to catch as many box as possible during the 28 sec session. The moving box randomly changes its direction but the velocity is the same within a session. The velocity increases 200, 400, 600 (pixel/sec) as the level of session increases. During the session, each subject performs three sessions with different complexity and self-load is measured. Figure 4(right) is a snap shot during this experiment.

4.2 Result and discussion

The initial self-load sl_0 measured before the experiment is distributed from 400 to 474. As we discussed in Section. 2, the initial self-load depends on the body shape and initial pose of the subject and it does not have importance. In the following two experiments, self-load SL is normalized, and replaced by $SL' = SL/sl_0$.

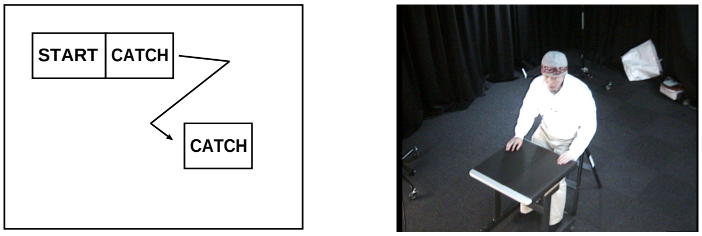
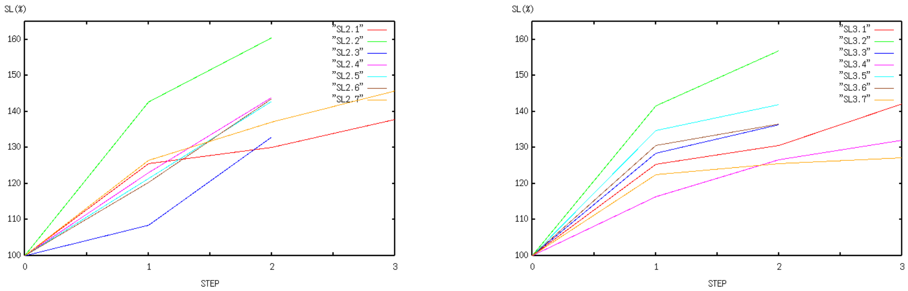


Fig. 4. Screen shown to the subject during the catch test (left) and a snap shot of the experiment (right)



(a) clap test

(b) catch test

Fig. 5. Comprexity of test and self-load $SL(\%)$

Figure. 5 (A) and (B) show the relationships of SL measure and complexity level of clap test and catch test respectively. Both of the figures show the strong correlation between SL and the complexity level, suggesting that SL can be used as the index of concentration.

In both tests, there are several subjects whose SL drops at level 3 complexity. However, it proved that some of them are moving their hand so fast and the system cannot recover the 3D volume due to the motion blur. The rest of them are those who gave up performing the requested task and not involved any more. So, those samples, which look conflicting with our estimation, do not conflict actually. Such samples are rejected from the Figures.

The difference between clap test and catch test is that catch test requires the subject continuous motion in proportion to the complexity. This implies contribution of kinetic energy to self-load increases as the complexity increases. However, we found that most of the dominant increase is posture maintaining energy. This suggests that concentration will appear as leaning forward pose. The result will match the previous work by Mota[6].

5 Conclusion

A feasibility study on observing human concentration with multiple cameras is described. A new measure, which is referred to as self-load, is proposed. Self-load is a energy consumption of the observed human keeping the same pose and motion. The pose is estimated from the 3D volumetric representation of the observed human and it is derived from 3D shape reconstruction technique like visual hull.

Through two scenarios of evaluation, we confirmed strong correlation between self-load and concentration. Also, we discovered leaning forward pose is appearing when a human is concentrating to a task. Currently, the situation in which we can measure self-load is limited due to the limitation of pose fitting and invisible force of the observed human. Never the less, it is a epoch that estimating human internal state with images is feasible.

Further study will include wider range of scenario to observe human concentration, estimation of human state other than concentration, and integration with other modality such as face recognition, sight line recognition, and speech recognition.

References

1. L. R. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993
2. Andrew Wilson, Aaron Bobick. "Learning Visual Behavior for Gesture Analysis". IEEE Trans. PAMI, vol. 21, no. 9, pp.884-900, 1999
3. T.B. Moeslund and E. Granum. "A Survey of Computer Vision-Based Human Motion Capture," CVIU, vol.81, no.3, pp.231–268, 2001
4. R. W. Picard and J. Healey, "Affective Wearables," Personal Technologies, Vol. 1, No. 4, pp. 231-240, 1997
5. R. Fernandez and R. W. Picard, "Signal Processing for Recognition of Human Frustration," Proceedings of ICASSP98. MIT MediaLag TR 447. 1997
6. S. Mota and R. W. Picard (2003), "Automated Posture Analysis for Detecting Learner's Interest Level." Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, CVPR HCI, June, 2003
7. Michael J. Lyons, Julien Budynek, and Shigeru Akamatsu, "Automatic Classification of Single Facial Images", IEEE Trans. PAMI, Vol;.21, no.12, pp.1357-1362, 1999
8. Seki Makito, Shimotani Mitsuo, Sumi Kazuhiko, Measurement of Driver's Facial Direction, proc. IMEKO XV World Congres, pp.51-56, June, 1999
9. A.V. Hill, "First and last experiments in muscle mechanics", London, Cambridge University Press, 1970
10. D.M.Gavrila and L.S.Davis. 3-D model-based tracking of humans in action: a multi-view approach. in Proc. IEEE Computer Vision and Pattern Recognition, San Francisco, 1996.
11. Takashi Matsuyama, X. Wu,Takeshi Takai, and Toshikazu Wada: Real-Time Dynamic 3D Object Shape Reconstruction and High-Fidelity Texture Mapping for 3D Video, IEEE Trans. on Circuits and Systems for Video Technology, Vol.CSVT-14, No.3, pp.357-369, 2004