Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-19 (2016) 367–372

Analysis of Human Pointing Behavior in Vision-based Pointing Interface System - difference of two typical pointing styles -

Kazuaki Kondo * **Genki Mizuno** * **Yuichi Nakamura** *

* *Academic Center for Computing and Media Studies, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan (email: kondo, mizuno, yuichi@ccm.media.kyoto-u.ac.jp)*

Abstract: This paper reports human pointing behaviors in vision-based pointing interface system, to make a mathematical model of them for designing easy-to-use interface. In natural pointing situations, we point targets at distant position with various postures, for example straight arm style or bent elbow style. We analyze their difference in pointing behaviors with assuming the pointing interface system as a feedback control model including an indicator. The difference had been confirmed in the step responses and the estimated parameters in the transfer functions, and matches to our actual experiences in those pointing styles. The estimation accuracy of indicated position from indicator's posture in the intermediate styles has been also analyzed. The results said that the reference point of indication smoothly moves from indicator's eye to his or her elbow according to the elbow joint angle.

Keywords: Pointing systems, Human-machine interface, Control theory, Closed-loop systems

1. INTRODUCTION

Nowadays we can get wide visual display devices inexpensively, which makes us easy to show various contents on large regions. This accelerates distance between display and users. In such environments, easy and intuitive remote interaction scheme is required rather than oscillatory interface like touch panel. For example, in presentation scene with slides on a wide screen, it is not reasonable for audiences to come close to the screen for pointing particular portion on the slide, directory. We usually use pointing gesture in such a case. The pointing interface system shown in Fig. 1 supports the remote pointing using on visual measurement. It shows a pointer at an estimated indication position on a display based on capturing indicator's posture. With this system, an indicated position becomes clear to the indicator and audience, which encourages smooth communication. We often use a raiser pointer for similar purpose, but it can be used only for indicating. The vision-based pointing interface controls displayed contents and materials according to user's posture, and will construct interactive scheme beyond mere indicating. Additionally when robots living with human and/or virtual agents connect to the pointing interface, these can get information about indicators and behave adaptively.

Kondo et al. (2015) tried to assume the vision-based pointing environment as a feedback loop model under classical control theory. It mathematically describes indicator's pointing behaviors and visual perception using transfer functions to simulate them with various configurations for designing easy-to-use interface, and to predict pointing behaviors for adaptive display. However the proposed model assumed an indicating posture with an arm being straight, and did not assume that with an elbow joint being bent that also appears in natural pointing situations. The purpose of this paper is to analyze human pointing behavior in the two pointing styles to expand the proposed pointing interface model to general one that can deal with arbitrary pointing types. We analyze how the pointing styles affect to pointing behaviors and how these can be described in the interface system model through measurement of actual pointing. Additionally what kind of method to estimate indicated positions should we select, in the case of the intermediate pointing style, e. g. slightly bent elbow postures are also investigated.

2. RELATED WORKS

The real-time vision-based pointing interface becomes practical with the progress of computational and visual sensing performance. However, we still have two considerable problems for natural pointing in daily environments. One is the difficulty in accurate measurement of pointing pose. Significant accuracy is required, especially for a target at a distance because even tiny errors on coordinates of body parts are amplified on a screen. We have several marker-less motion capture techniques based on visual sensing that do not interfere with an indicator's behavior like Shotton et al. (2011); Yoshimoto and Nakamura (2015), but the performance of those methods does not satisfy the requirement, because visual sensing is unstable to illumination change and occlusion. The latency arising from the sampling time of visual sensing and processing time for estimation also cannot be ignored.

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.10.593

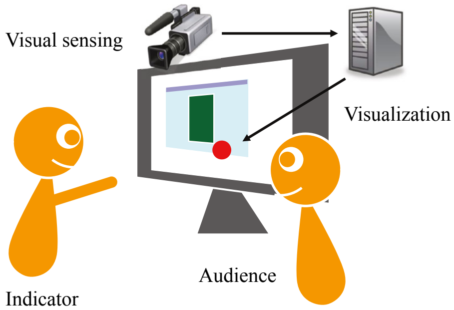

Fig. 1. An overview of a vision-based pointing interface. The vision sensor measures indicator's posture and estimates indicated position to display a pointer at that location.

The other is ambiguity in human pointing behavior. It depends on position of the indicating target, intention in indicator, and also individuality. Pointing pose relative to a screen indicates approximately where a person is pointing. Fukumoto et al. (1994) reported that a target position is on a line defined by a fingertip and a reference point inside of an indicator's body. The reference point moves depending on the pointing pose. As shown in Fig. 2, it is placed at an eye position for a distant target, but at an elbow position for a relatively near target. In addition, a geometric environment perceived by a human may not match an actual one. Knowledge of the relationship between a pointing posture and the indicated position is complicated and influenced substantially by various conditions. This makes it difficult to estimate indicated position from pointing pose accurately.

Characteristics of a transient pointing behavior during an indicator changes a pointing target have been analyzed for a long time. R.S.Woodworth (1899) proposed a pointing action model with a combination of feed-forward motions for rapid approach to a target position followed by feedback adjustments. Fitts (1954) reported that pointing duration increases with a larger moving distance or a smaller target, a notion known as "Fitts's law", which has been used in many studies because of its accurate approximation in various conditions.

The above problems suggest a need for additional schemes that reduce the influence of the ambiguity of pointing behaviors and measurement error to construct an easy-to-use pointing interface. Most of the conventional methods tackling this issue focus on how to visualize a pointer and contents on a screen. McGuffin and Balakrishnan (2005) proposed zooming the region around the pointer. This means that the target size and the distance from it feel larger in its neighborhood. To obtain a similar effect, controlling cursor size or cursor speed has also been proposed by Worden et al. (1997); Grossman and Balakrishnan (2005); Blanch et al. (2004). Retaining the pointer trajectory within the last short duration proposed by Baudisch et al. (2003) helps to recognize and predict the behaviors of the pointing interface intuitively. However, those methods assume a mouse interface and have not been evaluated with a remote vision-based pointing interface, as we assume here. Thus, we need a general framework that enables us to evaluate, compare, combine, and improve such conventional methods under various conditions.

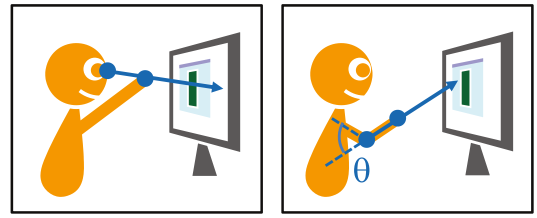

Fig. 2. The indicated position model for two typical pointing styles. (Left) with straight arm (Right) with bent elbow

3. MODEL OF POINTING INTERFACE SYSTEM

In this paper, We assume a vision-based pointing system of the sort in Fig. 1. It proceeds as follows.

(1) The indicator has the target he wants to indicate in his or her intention. We define its position as a reference position p_t.
(2) The computer estimates its location p_e based on visual sensing through the camera. The pointer is displayed at the location p_c conducted from p_e with some filters for visualization.
(3) The indicator recognizes the pointer location as p_r and adjusts his pointing posture to move it close to the target.
(4) Steps (2)-(4) continue until $p_t - p_r$ becomes 0. [1]

This procedure can be modeled as a feedback control loop, as shown in Fig. 3. The control model is constructed by H_g for the indicator's body kinematics, H_p for his or her visual perception catching a pointer, H_s for a computer estimating indicated position, and H_v for visualization filter. In this model, the indicator works as a controller with H_g and adjusts a feedback gain with H_p, simultaneously. The visualization H_v includes the pointer's shape, position, and so on. The two problems in a vision-based pointing interface, the pointing pose ambiguity and the error of pointing pose estimation, are in H_g and the noise d_s, respectively.

3.1 Pointing interface part

The estimation of indicated position H_s consists of visual sensing via cameras and a pose estimation algorithm based on the measurement. We assume that it correctly estimates indicated position with particular latency τ_s and the estimation error included in the noise term d_s, to formulate H_s as

$$H_s(s) = e^{-\tau_s s}. \tag{1}$$

On the other hand, the formulation of H_v is determined by the visualization methods of a pointer and contents. We assume to display a sufficiently small circle pointer at a smoothed position of the estimated pointing position during the latest short duration N. This visualization formulates H_v as

$$H_v(s) = \frac{1}{N} \sum_{n=0}^{N-1} e^{-(n\tau_s + \tau_v)s}, \tag{2}$$

[1] In this paper, we focus on only the indicator's behavior with assumption of audience perception being same as him.