MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN

Worker Behavior and Intension Modeling in Production Process

Takuma Funahashi Takayuki Hoshino Naoya Tokuda Takayuki Fujiwara Hiroyasu Koshimizu
Chukyo University
101, Tokodachi, Kaizu-cho, Toyota, 470-0393, JAPAN
t2784@sist.chuyko-u.ac.jp

Abstract

This paper proposes a vision-based human modeling by using both analytical property of human body performance and the concurrent gaze pattern for production process. Authors developed two systems required for 3D motion analysis: Firefly Capturing Camera (FCC) and gaze analysis: Passive Eye Camera (PEC). Moreover, FCC, PEC and other capture systems measured veteran and beginner worker's motion and gaze pattern for analyze worker's skill especially vision-based. Authors demonstrated these systems are effective to construct the vision-based model of human behavior in production process.

1. Introduction

In the manufacturing of Japan, the deindustrialization by the overseas transfer of production basis gives several negative effects to the industrial accumulation region. Especially, expert skill transfer is strongly limited in foundation of technological industry. Moreover, the coming aging society and the decrease of expert successors cause the lack of succession of a basic technology, expert skill and creations of knowledge.

1.1. Precedent researches (a brief survey)

Watanuki et al.[1] proposed a virtual reality based knowledge acquisition and job training system for casting design, which is composed of the explicit and tacit knowledge transfer systems using synchronized multimedia and the knowledge internalization system using portable virtual environment.

J-Y. Fourquet et al. study is to provide designers with a realistic prediction of human movements in order to evaluate tasks and workplaces from the ergonomic viewpoint [2]. These studies proposed motion capture based e-learning system. However, model of human behavior especially production process was not realized.

1.2. Proposed method

Authors propose a new vision-based paradigm for analyzing and modeling human performance and gaze pattern in production process. Seko et al. developed two systems needed for 3D motion analysis and modeling. Firefly Capturing Camera (FCC)[3] uses the monocular camera with large spherical aberration of lens to measure 3D positions of point light sources attached on an object in real time without any special lighting. Passive Eye Camera (PEC)[4] was implemented based on the facial image processing such as Hough transform for iris detection and gaze analysis. Authors proposed the concurrent utilization of FCC and PEC for detecting human body performance and modeling in the production process. Moreover, FCC, PEC and other systems measured veteran and beginner worker's motion and gaze pattern for analyze worker's skill especially vision-based.

2. System Configuration

2.1. FCC and multi cameras

3D positions of light sources are successively calculated by the measurement by using FCC. By this method authors realized the motion capturing of multiple LED's arranged in space. Six LED's with wavelength of 950 nm were set on a card with a battery as shown in Fig 1. Authors used the same camera as mentioned above and a visible light cut filter in front of the lens to remove environment light. Video image was input to a desktop PC with3.4GHz, 3D positions of LED's were calculated, and their 3D positions with the spheres were shown in 3D graphics at the frame rate of 30 fps. The visible area of FCC is shown in Fig. 2.

Authors could experimentally demonstrate the accuracy and speed of the LED's motion that was moved in various manners by hand within this visible area.

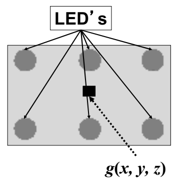


Figure 1. LED's on card (Marker)

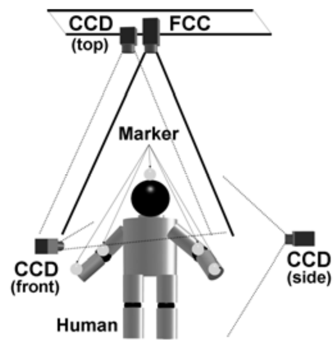


Figure 2. Visible area of FCC and Multi Cameras

Human performance is measured by one FCC and 3 different CCD cameras. FCC can detect the 3D positions of head, left and right hands and object in real time processing (Eq.1). CCD cameras capture the top, front and side views of images of human inspection performance (Eq.2).

These data is very important to analyze the human motion. FCC measures precisely the 3D motion how the human moves the head, hand and arm. Multi cameras verify the motion captured by FCC and also provide color images for image processing.

$$(x_i, y_i, z_i)^{(k)} \quad \text{(position } k = 1(head), 2(lh), 3(rh), 4(obj)) \quad \text{(sequence } i = 1, 2, 3, \ldots\ldots) \tag{1}$$

$$f_i^{(front)}, f_i^{(side)}, f_i^{(top)} \quad \text{(sequence } i = 1, 2, 3, \ldots\ldots) \tag{2}$$

2.2. PEC

The PEC system is composed by the devices with simple specification as follows;

(1) CPU : 1.8GHz
(2) RAM : 1.0GB
(3) OS : Windows XP
(4) Software : Visual C++
(5) Camera : 0.3Mpixel CMOS Sensor

The system is shown in Figure.3. This system obtains the image from CCD camera, the face region extraction procedure is applied to the image, and the iris is recognized in the face area. Gaze points are acquired by using the locus of the iris centre given by the iris detection. And another synchronized CCD camera captures the object image just in the same manner of human's subjective view (Eq.3).

These image data is analyzed for extracting the intentional gaze of the human by calculating the correlation among gaze, object and head motions (Eq.3). Gaze, saccade and spatial distribution are measured by using the eye mark pattern. And object pose is simultaneously analyzed for modeling the best gray contrast feature, sufficient resolution and suitable lighting for the acquired image by checking the subjectivity image at the synchronized cameras.

$$f_i^{(face)}, f_i^{(obj)} \text{correlation } (x_i, y_i, z_i)^{(4)} \quad \text{(sequence } i = 1, 2, 3, \ldots\ldots) \tag{3}$$

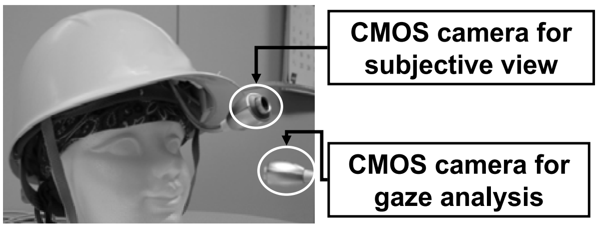


Figure 3. Overview of PEC

3. Multi-Modal Image Sensing and Analysis

3.1. Comparing movie contents with inspection intention

In visual inspection, we cannot know the detailed intention what the inspector wants to do. Then, authors asked the inspector to verbalize his intention at the every moment in the inspection process. Images of the multi cameras can support to translate the internal intention into verbal information (Eq. 4) by super-imposing the texts into the images.

$$Intention_i = f_i^{(front)}, f_i^{(side)}, f_i^{(top)} + Cap_i \quad \text{(sequence } i = 1, 2, 3, \ldots\ldots) \quad (Cap: \text{super - impose}) \tag{4}$$

3.2. Comparing motion with inspection intention

In addition, we cannot understand precisely how the inspector performs in the 3D space according to his inspection intention. For example, when he raises his elbow, this motion might mean the critical factor of inspection, and might be just a behavior of habit. Therefore we should compare directly the 3D motion with inspection intention.

3D spatial positions were measured by FCC and the verbal descriptions of the inspection intention are compared with the 3D motion data (Eq. 5).

$$Intention_i = (x_i, y_i, z_i)^{(k)} + Cap_i \quad \text{(position } k = 1(head), 2(lh), 3(rh), 4(obj)) \quad \text{(sequence } i = 1, 2, 3, \ldots\ldots) \quad (Cap: \text{super - impose}) \tag{5}$$

3.3. Cognitive hypothesis of inspection

Intentional inspection process was visually exposed in the previous sections. Then, we should investigate thoroughly the relations among body parts. "Eq.6" correlates head position with both arms, and "Eq. 7" correlates right arm with left arm. We can know these cooperative motions give us some hypothesis of inspection based on the computer vision. "Eq. 8" correlates head position with the object, and "Eq.9" correlates the object with both arms. From these hypotheses extracted from the visual information, we can know the best positional relations among cameras, feasible resolution and expected contrast of the image for image processing, and lighting condition for automated inspection.

$$(x_i, y_i, z_i)^{(1)} \text{correlation} \left\{ (x_i, y_i, z_i)^{(2)} + (x_i, y_i, z_i)^{(3)} \right\} / 2 \tag{6}$$

$$(x_i, y_i, z_i)^{(2)} \text{correlation} (x_i, y_i, z_i)^{(3)} \tag{7}$$

$$(x_i, y_i, z_i)^{(1)} \text{correlation} (x_i, y_i, z_i)^{(4)} \tag{8}$$

$$(x_i, y_i, z_i)^{(4)} \text{correlation} \left\{ (x_i, y_i, z_i)^{(2)} + (x_i, y_i, z_i)^{(3)} \right\} / 2 \tag{9}$$

4. Accuracy of PEC and FCC

4.1. Experiment of PEC

Experiment was executed as follows:

Stimuli image designed as 38x38mm/matrix with 9x7 matrixes was displayed on the monitor, PEC was attached at the head, and a person looks at the monitor. The distance between person and monitor were 650mm, and the size of monitor was 340mm(w)×280mm(H). Authors investigated the view resolution by analyzed gaze point and eye movement measurement.

The eye movement range detected from the center of iris position was 7 pixel (MAX:27 MIN:20) in vertical direction, and 19 pixel (MAX:70 MIN:51) in horizontal direction. Here, the viewing angle can cover the monitor in Figure. 4 and the parameters became $\theta_{vertical}$ 24.2° in vertical angle and $\theta_{horizontal}$ 28.4° in horizontal angle, PEC resolution power became 3.45° in a vertical angle and 1.5° in the horizontal angle.

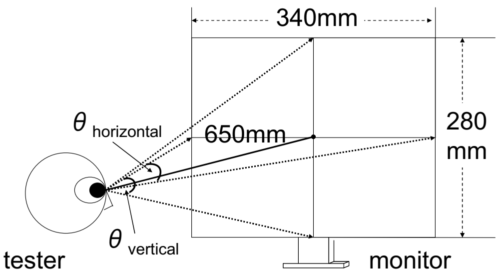


Figure 4. Parameters for the experiment viewing among monitor and human.

4.2. Experiment of FCC

Worker performed all inspection operations in the environment designed by Figure. 1 with attached 5 LED markers. Authors calculated the average number of appearance LED marker and the frequency of the average number of the detected LED marker. The result is shown in Table 1.

In the result of six time trials, the average number of appearance LED marker was 3.55 and the frequency of the average number of detected LED marker were as follows: 1; 0.4%, 2; 10.55%, 3; 35.45%, 4; 40.95%, 5; 12.7%.

4.3. Considerations

In the FCC result, the appearance frequency of FCC markers was lower in precision than 3D Motion Capture system (Motion Analysis, Inc.), and it was known FCC marker was frequently suffered from the occlusion. However, workers head and arm motion could be captured in real time by subjective estimation.

In the PEC result, PEC resolution power of 1.5 degree was not precise enough compared with Eye Mark Recorder (nac Image Technology., Inc.) resolution power of 0.1 degree, it is coarsely known from these eye gaze patterns that the worker is looking intensively at the body of object.

5. Worker's Motion Measurement

5.1. Measurement environment

The measurement environment is shown Figure. 5 and 6.

Authors measured Veteran (with about 40 year carrier) and Beginner (with about 1 month carrier), where each person performed arbitrary working motion by 8 times to one object. Measurement work space is illuminated by only indirect lighting, and the used motion capture and eye mark recorder were as follow;

(1)Motion Capture System

- MAC3D SYSTEM hawk-i : (MotionAnalysis,. Inc.)
- AC3D SYSTEM Eagle : (MotionAnalysis,. Inc.)

(2)Eye Mark Recorder

- EMR-8B : (nac Image Technology,. Inc.)

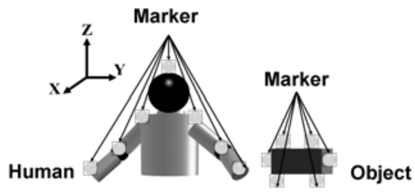


Figure. 5. Position of attachment marker

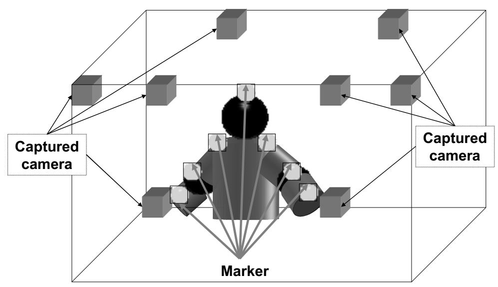


Figure. 6. Overview of measurement environment

Table. 1. Result of FCC's motion capture

Total Frame		2457		2747		2658		2789		3098		3003	
Average sum of LED		3.22		4.02		3.87		3.60		3.61		3.51	
LED appearance frequency	1	0	0.0%	0	0.0%	0	0.0%	0	0.0%	46	1.5%	19	0.6%
	2	294	12.0%	12	0.4%	50	1.9%	122	4.4%	79	2.5%	289	9.6%
	3	1330	54.1%	633	23.0%	653	24.6%	1240	44.4%	1261	40.7%	1152	38.3%
	4	834	33.9%	1389	50.5%	1554	58.4%	1050	37.6%	1373	44.3%	1216	40.5%
	5	0	0.0%	714	26.0%	402	15.1%	378	13.5%	340	11.0%	328	10.9%

5.2. Result of distance between head and object

In this paper, authors paid attention to the position between head and object as shown Figure. 7, calculated the difference between the center of head Z coordinate value and center of object Z coordinate value by "Eq. 10". These results were shown in Figure 8(a)-(h) each worker.

$$\theta_g = \tan^{-1}\left(\frac{\mathrm{EyePoint}(y_n) - \mathrm{EyeCentre}(y_n)}{\mathrm{EyePoint}(z_n) - \mathrm{EyeCentre}(z_n)}\right) \times 180 \div \pi$$
$$\theta_w = \tan^{-1}\left(\frac{\mathrm{M3}(y_n) - \mathrm{M6}(y_n)}{\mathrm{M3}(z_n) - \mathrm{M6}(z_n)}\right) \times 180 \div \pi \qquad (10)$$

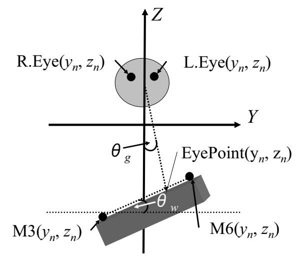


Figure. 7. Definition of parameters between head and object

5.3. Consideration

As for the distance between head and object, veteran is very smooth and stable in motion as given in the result of Figure. 8(a) and (c). On the other hand, beginner is rough and unstable in motion as given in the result of Figure. 8(b) and (f). Veteran could keep a constant distance between head and object compared with Beginner. Head angle and object angle were also in the similar property.

The object angle is in the similar motion for both of workers. However, the fact that the difference of head angle motion and the distance between head and object were very similar was originated from the reason why veteran's motion were kept for saving the contrast change by indirect lighting.

6. Conclusion

In this paper, Authors proposed a total paradigm of vision-based analytical system for modeling human performance and gaze behavior especially in industrial production process. Some of the proposed motion capture system provided the successful results in the preliminary demonstration, and the analytical investigations on the workers performances were promising for constructing the vision-based model of human behavior in production process.

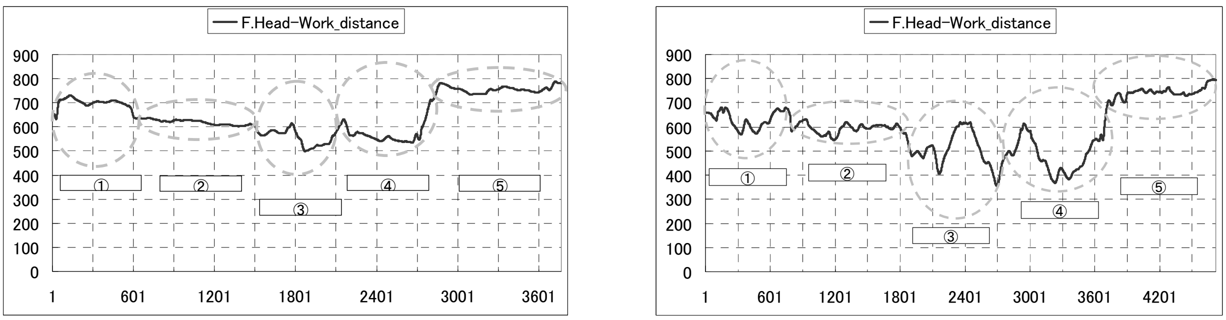
References

[1] K. Watanuki and K. Kojima, "Knowledge Acquisition and Job Training for Advanced Technical Skills Using Immersive Virtual Environment", Journal of Advanced Mechanical Design, Systems, and Manufacturing, Vol. 1, No. 1, pp.48-57. (2007)

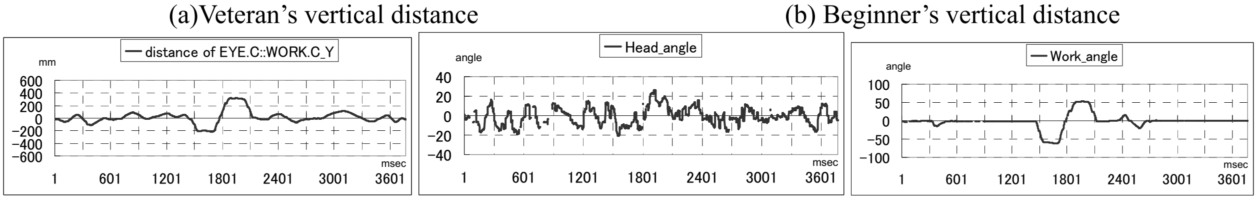
[2] J-Y. Fourquet, V. Hue and P. Chiron, "OLARGE : on kinematic schemes and regularization for automatic generation of human motion and ergonomic evaluation of workplaces", Proc. Of 33rd Annual conference of the IEEE Industrial Electronics Society, PF-013498. (2007)

[3] Y. Seko, K. Murai, H. Hotta and J. Miyazaki, "A new method of measure 3D positionof a light source by tracking the ring images made by a hemispherical lens", 2003 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 2003, pp.409-412 (2003)

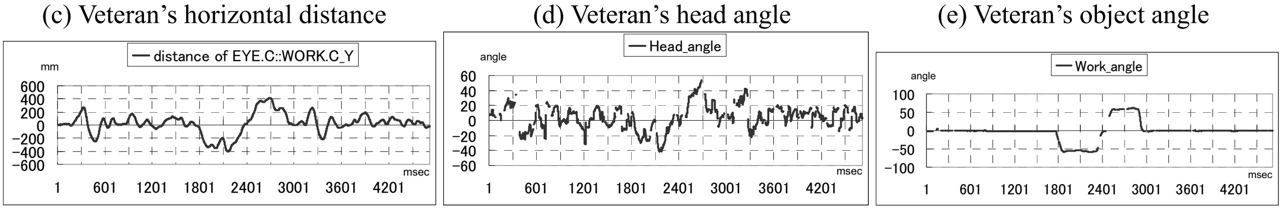
[4] T. Funahashi, T. Fujiwara and H. Koshimizu, "Facial Media for Non-verbal Communication in Production Processes", Proc. Of 32nd Annual conference of the IEEE Industrial Electronics Society, SS22:PF-015474. (2006)



(a)Veteran's vertical distance (b) Beginner's vertical distance



(c) Veteran's horizontal distance (d) Veteran's head angle (e) Veteran's object angle



(f) Beginner's horizontal distance (g) Beginner's head angle (h) Beginner's object angle

Figure 8. Result of distance between head and object