Real-time Evolutionary Recognition of Human with Complex Background and Varing Illumination

Jianing ZHU[1], Mamoru MINAMI[2]

[1] *Fukui University, Fukui, Japan, jianing4780@hotmail.com*
[2] *Fukui University, Fukui, Japan*

Abstract

Many trials to realize human's visual recognition by image processing have been carried out. Recognition and tracking ability of human are expected to be applied for a security system and a field of ITS (Intelligent Transport System). This paper aims at construction of real-time recognition system with robustness against changing of illumination. We employ Genetic Algorithm (GA) and a gray-scale image termed here as raw-image to execute recognition process by a model-based matching method in real time. Proposed recognition method can be divided into two main functions, that is, global and local searching, which is switched depending on the matching degree of a human in the raw-image and a model to be matched to the human. Furthermore, we improved the recognition performances, to shorten recognition time and to raise the accuracy, by adding gazing operation in local searching, which is inspired from gazing action of human. Moreover, in order to improve the recognition system to be robust against lighting condition varieties, an Illuminance-depending Gazing GA, which changes the threshold value automatically based on the brightness value of the image was implemented to the system toghther with an evaluation of the skin color while keeping the real-time nature.

1 Introduction

So far, many trials to realize human's visual recognition by image processing have been carried out. Human recognition and tracking are expected to be applied for a security system of buildings and a field of ITS (Intelligent Transport System) to avoid traffic accident harming humans. For example, in a field of ITS, an image recognition systems are mounted on a car to support human's visual recognition process and, if possible, to replace it for automatic driving system whose technology consists of lane recognition, obstacle detection, road signs detection, walker recognition, and so on. This research aims at construction of recognition system of human in real time. Here, the real-time recognition means that an object should be recognized in the images input within video rate, i.e., 33 [ms].

For image recognition purposes, many researchers [1]-[4] have employed a binary image and some edge extraction methods that require several preprocessing steps. The various filtering stages seem to be a time-consuming process, and to us, these are not convenient for real-time recognition. The tradition image recognition of human [9] often uses extraction technology of many characteristics for shape or position relation of eye, mouth and nose. however, there are some problems to extract a lot of characteristics because of non-clear shadow, which makes image recognition of human difficult. Then, we use directly the unprocessed gray-scale image termed here as raw-image. Basically, this research is based on a model-based pattern-matching method. We employ a model designated as surface-strips model for the recognition purposes of a human considered here as the target. An objective function, whose computation is based on the configuration of the surface-strips model, is used to evaluate the extent to which the surface-strips model matches with the object being imaged, by changing the recognition problem into an optimization problem. Therefore, We use a Genetic Algorithm (GA) in the image recognition, because of its high performance of optimization.

GA is well known as a method for solving search and parameter optimization problems [5]. Moreover, to use the GA process in real time, i.e., to extract its position from the consecutively input images, we used the GA such a way that every input image is evaluated only one time by target-model-based fitness function, which we named Step GA [6]. Furthermore, in order to increase the tracking performance to a target human, here we employ hybrid-searching method, which is a localized search technique of a GA combined with a global GA process, that is, conventional GA without loosing real-time nature of Step GA. Using this localized search technique, we confirmed in previous researches that the

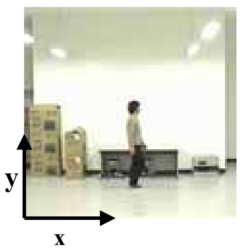

Figure 1: *Raw Image*

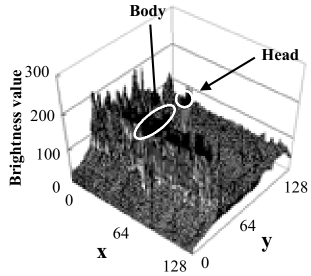

Figure 2: *Brightness*

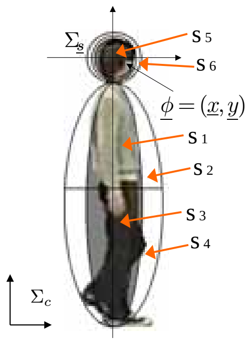

Figure 3: *Serching model*

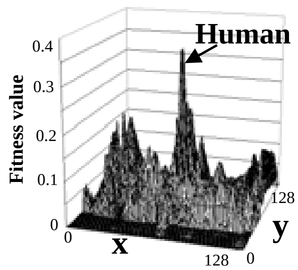

Figure 4: $F(\underline{\boldsymbol{\phi}})$

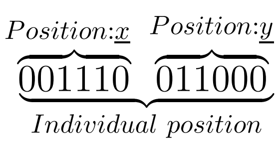

Figure 5: *Structure of GA's gene*

hand-eye manipulator can catch a fish swimming in a pool with a net attached at the hand [7].

To make our recognition process to be robust against illuminance changing of the environment while keeping the tracking performance of the Step-GA, the switching of the local search process should be determined depending on illuminance level of the input image. In this report, in order to perform the recognition of a human in spite of changing of the lighting condition, we propose an illuminance-depending gazing GA [8], which changes the threshold value to determine the gazing area automatically based on the brightness level of the image. Moreover, a new real-time recognition system including the information of skin color of human is supposed to improve the tracking performance of walking human. We performed experiments to recognize a human to confirm the effectiveness of our proposed recognition system.

2 Evolutionary Recognition

2.1 Real-time Recognition by Gazing Step GA

Consider the 2-D raw-image of a target human shown in Fig.1, its corresponding 3-D plot is shown in Fig.2. In this figure, the vertical axis represents the image brightness values, and the horizontal axis, the image plane. To search for a such target human in the raw-image, a geometrical human model as shown in Fig.3 is used. In this research, a geometrical model, which is composed of an internal surface and contour-strips, is employed. The internal surface approximated the most the 2-D top surface of the target. Such model is designated as surface-strips model. In Fig.3, S_1, S_3, S_5 denotes the inside surface of the model, and S_2, S_4, S_6 denotes the contour-strips. Also, the combination is designated as S. When the position of surface-strips model S is defined as a function of $\underline{\boldsymbol{\phi}} = [\underline{x}, \underline{y}]^T$, which designates the position of the origin of the model, then S moves in the camera frame and a set of x-y coordinates in the moving model is expressed as $S(\underline{\boldsymbol{\phi}})$. And moreover, the brightness distribution of raw-image corresponds to the area of the moving model is expressed as $p(\tilde{\boldsymbol{r}}), \tilde{\boldsymbol{r}} \in S(\underline{\boldsymbol{\phi}})$, then the evaluation function $F(\underline{\boldsymbol{\phi}})$ representing matching degree of the model and the raw-image, of the moving surface-strips model is given as follows in Eq.(1).

$$F(\underline{\phi}) = \sum_{i=1,3,5} \Big| \sum_{\tilde{\boldsymbol{r}} \in \underline{S}_i(\underline{\boldsymbol{\phi}})} p(\tilde{\boldsymbol{r}}) - \sum_{\tilde{\boldsymbol{r}} \in \underline{S}_{i+1}(\underline{\boldsymbol{\phi}})} p(\tilde{\boldsymbol{r}}) \Big| \quad (1)$$

This function means the integrated brightness difference of the input raw-image between the one of the internal surface and the one of the contour-strips of the surface-strips model. The filtering result by using Eq.(1) of the surface-strips model-based fitness function, with respect to Fig.1 is shown in Fig.4. The filtering result in Fig.4 has a peak corresponding to the target human in the raw-image. An evaluation using the surface-strips model means that $F(\underline{\boldsymbol{\phi}})$ takes into account the differentiation between an object signal and the background brightness. As long as we can set such an environment that the highest value of $F(\underline{\boldsymbol{\phi}})$ is obtained only if S fits to the target object being imaged. Then the problem of recognition of a human and detection of its position is converted to a searching problem of $\underline{\boldsymbol{\phi}}$ such that maximize $F(\underline{\boldsymbol{\phi}})$. $F(\underline{\boldsymbol{\phi}})$ is used as a fitness function of GA.

GA is well known as a parallel search and parameters optimization algorithm. The GA is viewed as an optimization method since the iterative evolution process of the potential solutions toward better solutions is equivalent to the process of optimizing the objective function used as a fitness function in GA search. The term "parallel", in "parallel search" above is related to the implicit parallelism of GA. Implicit parallelism is explained in Goldberg (1989, pp. 40-41) [5]. The GA operates with a population of searching binary strings designated as individuals, shown in Fig.5, considered to be the potential solutions to

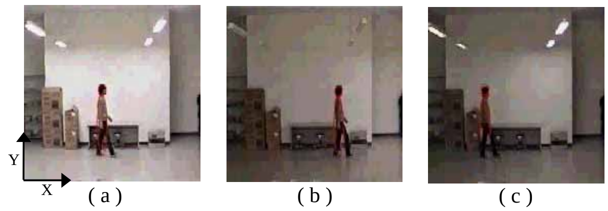

Figure 6: *3 Input Images*

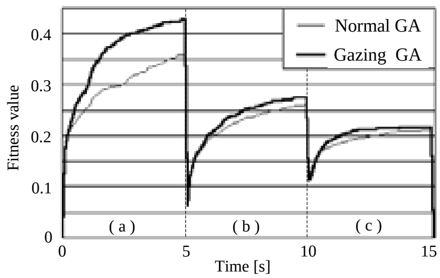

Figure 7: *Recognition Result of Normal GA Search and Gazing GA Search*

a given problem. The search by a GA for the solution is performed through an evolution process, from generation to generation.

To recognize a target in a dynamic image, the recognition system must have real-time nature, that is, the searching model must converge to a human in the successively input raw images. We have proposed a new idea of an evolutionary recognition process for dynamic image, in which the GA is applied only one time to the newly input raw image. Therefore every input image is evaluated only one time, we named it as "Step-GA" [6].

3 Recognition Experiment

3.1 Convergence Speed and Adaptation for Illuminance Changing

Experiments of recognition and tracking of a human in indoor environment using raw images have been performed to appraise the performances of the proposed scene recognition method. First, recognition experiment has executed using three fixed images of different illuminances as shown in Fig.6. In this experiment, the images have been input into the recognition system at 0, 5 and 10 [s], as a step-input manner. Moreover, this experiment has been repeated for 100 times. Fig.7 shows the mean value of the best fitness values of both normal GA and gazing GA versus the time used for the recognition, for comparison between the results of the normal GA (Conventional GA) and the results of the gazing GA (illuminance-depending). For all images, the fitness value of the Illuminance-depending Gazing GA search method increases quickly and higher than the one of the normal GA. These results indicate, at first, it can shorten the time for the recognition, that means proposed system improves the dynamical response of the recognition, relating to the improvement of the tracking performance to moving object. The second is the improvement of the recognition accuracy pointed out by the higher fitness value, which means closer matching of the searching model to the target human in the input images. These results show the effectiveness of the proposed search technique for real-time recognition having an adaptation for illuminance changing of the environment.

3.2 Real-time Recognition of Human in Dynamic Scene

Furthermore, we would like to confirm the validness of the proposed method while using dynamic scene. In the real-time tracking experiment, the images are obtained from a camera fixed in a room. This experiment has been performed at night, and we made a situation being changed the illuminance by putting on and off the light intentionally. Figure 8 presents a sequence of the tracking results observed by the CCD camera, when a human is walking. The white and connected two circle is representing a position of the detected human. Here, we change the searching area of gazing GA with 3 steps (level 1,2,3), and the calculation time is 30 [ms] for Step-GA recognition, by the computer DELL optiplex GX1 (CPU:Pentium2, 400MHz). The experimental results have confirmed the effectiveness of the proposed method by showing the robustness against an illuminance changing in indoor environment.

3.3 Human Recognition Experiment with Complicated Background

In this section, the experiment is performed under the complicated background environment. Figure10 describes the recognition result using the static scene of Fig.9. Because the fitness of the position where a person exists in does not show the maximum in input images, it cannot recognize human by traditional technique in Fig.10.

As for the traditional technique, human is characterized by the difference of brightness value between human domain and background, which is calculated by eq.(1), but TV is more similar with the shape of head so that it was misrecognized under the complicated background environment shown in Fig.9. So the next section suggests the method that can recognize human even under the complicated environment and verified by experiment.

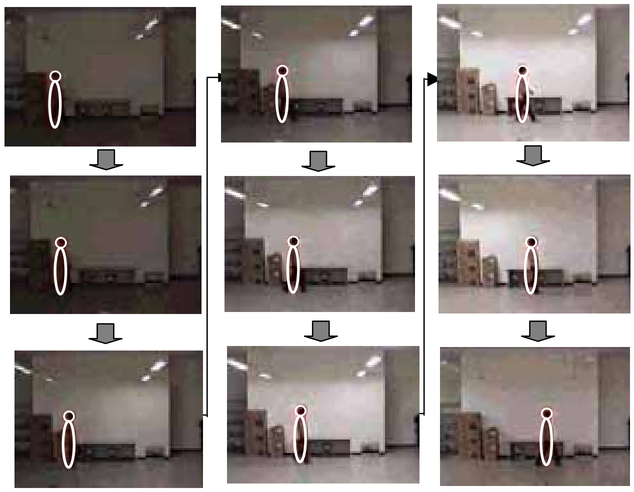

Figure 8: Recognition Result of Indoor Environment

Figure 9: Input Image

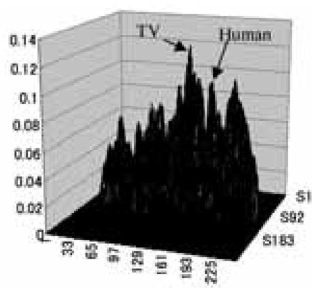

Figure 10: $F(\underline{\phi})$ Distribution

4 Recognition Using Color of Face

4.1 Effectiveness of HSV Color System

In this chapter, the recognition method paying attention to the color of human skin is supposed. Regarding human as the objective of recognition, the effectiveness of recognition can be improved while adding extraction of color of the skin which is one of human characteristic to the evaluation function. However, the color of everyone's skin is different. The sample image and RGB value of human skin are shown as Fig.11. Sample 1-8 stand for the skin color of Japanese, sample 9 is the white's, and sample 10 is the black's. The 1-8 are sample about one person under different situations such as outdoors, indoor, illumination is high or low etc. From the value of RGB, it is understood that there is great difference among the RGB value at the various situation even the same human.

So that, the HSV color system is considered (Here, the conversion expressions from RGB value to HSV value are omitted). H[0-359] of HSV represents the hue, which defines the direction of 3:00 at the hue circle shown in Fig.12 as 0, and the value becomes big counterclockwise. S[0-1] represents color, which is expressed by the radius of hue circle. V[0-255] is luminosity, and is the length of axis vertical to the hue circle. Color becomes dark with the value of V decreasing. Here, the value of H shown in Fig.11 should be noticed. Though the range of H is 0-359, the color of skin can be specified by limiting the value of H to 0-30 without relation to the change of illumination and race. Fig.13 describes skin pixels plotted in HS space. It is easy to understand that the color of skin can be limited only by H from Fig.13.

Sample	1	2	3	4	5	6	7	8	9	10
R	180	208	141	173	159	184	136	206	246	105
G	139	145	96	120	101	125	89	171	223	52
B	119	130	77	112	94	121	61	151	217	44
H	19.7	11.5	17.8	7.9	6.5	3.8	22.4	21.8	12.4	7.9
S	0.34	0.38	0.45	0.35	0.40	0.34	0.55	0.27	0.12	0.58
V	180	208	141	173	159	184	136	206	246	105

Figure 11: RGB and HSV

Figure 12: Hue Circle

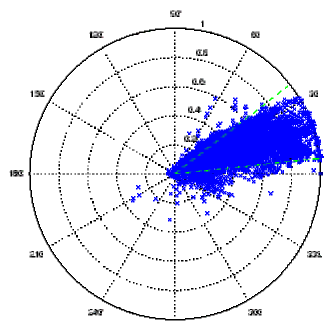

Figure 13: Skin pixels plotted in HS space

Then it is tried to extract the skin color by utilizing the value of H as Fig.9, Fig.14 shows the result. The pixels in range of 0-30 are expressed by yellow green. By Fig.14, it can be identified that the skin color of face and arm can be extracted, however, a wooden bookshelf with the same color of skin was also be recognized..

4.2 Recognition using Skin Color

It takes time to convert into HSV color system as preprocessing at all input images range, which is against the main premise of this study that is the nature of real-time. In this report, it does not do image conversion at all input image range as preprocessing, but as added value of conventional model filtering, a clause to evaluate the color information of skin is

Figure 14: *Result of Skin Recognition using H[0-30]*

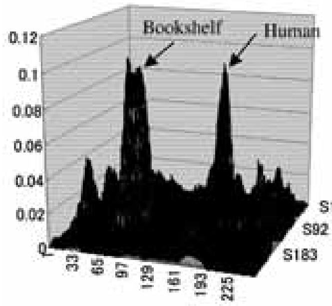

Figure 15: *Filtered Image using H Value*

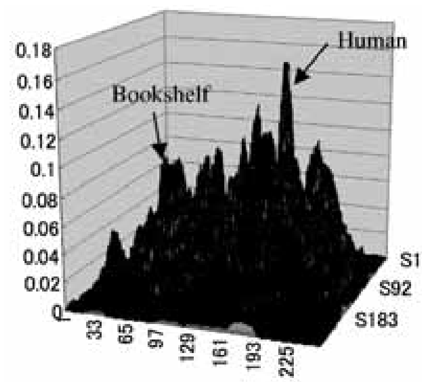

Figure 16: *Result of New Recognition Method*

added to the fitness function eq.(1), and the target is recognized directly from the input image without any conversion to shorten the conversion time to zero. In this study, the area of skin color underlying the searching model of S_5 shown in Fig.3 that represents the head part of human. Eq.(2) shows a new evaluation function of human recognition which includes both the result of model filtering of eq.(1) and the color information of face skin.

$$F_{HSV}(\underline{\phi}) = \sum_{i=1,3,5} \Big| \sum_{\tilde{\boldsymbol{r}} \in \underline{S}_i} p(\tilde{\boldsymbol{r}}) - \sum_{\tilde{\boldsymbol{r}} \in \underline{S}_{i+1}} p(\tilde{\boldsymbol{r}}) \Big| + \sum_{\boldsymbol{r} \in \Omega \cap S_5} h(\boldsymbol{r}) \quad (2)$$

The first term in the right side of function is to evaluate the human shape above-mentioned, and the second is to evaluate color information of face. The second term is used to convert RGB value of face into HSV value, and it is evaluated as skin color of human in the case that the value of H belongs to 0-30. The process evaluating color information of the face is described as the followings.

At first, the color infomation(RGB) of face is set as $\boldsymbol{p}$ and HSV data is set as $\boldsymbol{q}$.

$$\boldsymbol{p}(x,y) = \begin{bmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{bmatrix} \quad (3)$$

$$\boldsymbol{q}(x,y) = g\big(\boldsymbol{p}(x,y)\big) \quad (4)$$

$$= \begin{bmatrix} q_H(x,y) \\ q_S(x,y) \\ q_V(x,y) \end{bmatrix} \quad (5)$$

Second, when Ω is defined as following, h of the second term in the right side of eq.(2) can be expressed as eq.(9).

$$\Omega = \big\{ \boldsymbol{r} = (x,y) \bigm| 0 < q_H(x,y) < 30 \big\} \quad (6)$$

$$h(\boldsymbol{r}) = \begin{cases} 1 \ (\boldsymbol{r} \in \Omega) \\ 0 \ (\boldsymbol{r} \notin \Omega) \end{cases} \quad (7)$$

Figures 15 and 16 show the recognition results that applied the above-mentioned new method for Fig.9. Figure15 is the distribution that is obtained by calculation $q_H(x,y)$ in eq.(5). As a result, evaluation concentrates on both the human face and bookshelf from Fig.14. Fig.16 is the compound result about the Fig.15 and Fig.10. It can be identified that human position can be precisely recognized. But when the walking human can be recognized or not depends on the converging ability of GA, that is, it can converge to the moving highest peak faster than the walking speed of human.

4.3 Recognition of Human in Dynamic Scene

Figure17 shows the result that method above is applied to dynamic scene. It was calculated with 31ms per generation of GA, and confirmed the effectiveness of human recognization in real time when there is complicated background environment and similar object with human.

5 Conclusion

We have proposed in this paper a vision related technique for a real-time recognition of a human, which utilizes the global search feature of a genetic algorithm (GA) together with a local search technique of the GA, and also the unprocessed gray-scale image called here as raw image. And moreover, in order to perform the recognition of a target human adapting to changing of the lighting environment, we propose an Illuminance-depending Gazing GA, which changes the threshold value to change local search area automatically based on the brightness

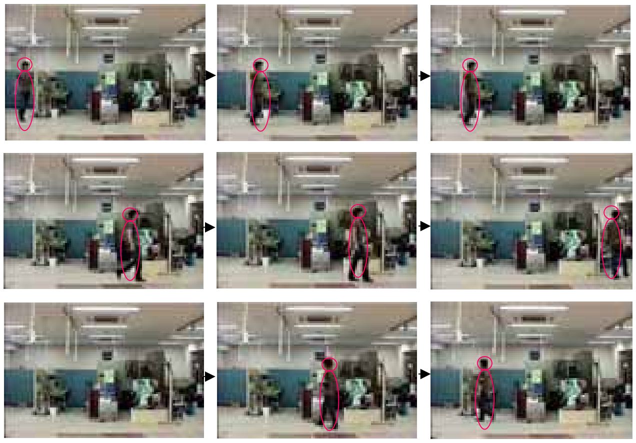

Figure 17: *Recognition Result of Real-Time*

value of the input image, i.e., illuminance, by using a relation that the threshold value is proportional to an illuminance of environment in the input image. Results of the experiment to track a target human whose position and the illuminance are changed by stepping manner, have shown the effectiveness of proposed GA search method. Furthermore, a new recognition method including the color information of human skin is supposed to improve performance of the human recognition. The experimental results confirmed the effectiveness of human recognization in real-time under the complicated background environment for the dynamic scene.

As the next experiment, we plan to further verify the effectiveness about the method suggested above by the recognition experiment of 100 humans. Furthermore, we will try to realize more robust recognization system of human.

References

[1] T. Fukuda and K. Shimojima "Intelligent Control for Robotics," Computational Intelligence, pp.202-215, 1995.

[2] G. Ao, H. Akazawa, M. Izumi and K. Fukunaga, "A Method of Model-Based Object Recognition," Japan/USA Symposium on Flexible Automation, vol.2, pp. 905-912, ASME, 1996.

[3] T. Nagata and H. Zha, "Recognizing and Locating a Known Object from Multiple Images," IEEE Transactions on Robotics and automation, vol. 7, No. 4, pp. 434-447, 1991.

[4] K. Sumi, M. Hashimoto and H. Okuda, "Three-level Broad-Edge Matching based Real-time Robot Vision," in Proc. IEEE Int. Conf. on Robotics and Automation, pp. 1416-1422, 1995.

[5] D.E.Goldberg, "Genetic algorithm in Search, Optimization and Machine Learning. Reading", Addison-Wealey, 1989

[6] M.Minami, J.Agbanhan and T.Asakura: "Manipulator Visual Servoing and Tracking of Fish Using Genetic Algorithm", Int. J. of Industrial Robot, 29-4, pp.278-289, 1999

[7] M.Minami, H.Suzuki, J.Agbanhan, T.Asakura: "Visual Servoing to Fish and Catching Using Global/Local GA Search", Int. Conf. on Advanced Intelligent Mechatronics, Proc., 2001, pp.183-188.

[8] M.Minami, H.Suzuki, M.Miura: "Real-Time Robust Recognition of Human using Illuminance-depending Gazing GA", Knowledge-Based Intelligent Information Engineering Systems and Allied Technologies, KES 2002, pp.1377-1382.

[9] ITO R, NAKAZAWA K, NAKAJIMA M: "Human Recognition Using 3D-gray-scale Face Image", Lect Notes Comput Sci, VOL.1451, PAGE.1005-1012 1998