Human Guide Tracking Using Combined Histogram of Oriented Gradient and Entropy Difference Minimization Algorithm for Camera Follower

Fitri Utaminingrum, M. Ali Fauzi, Yuita Arum Sari, Sigit Adinugroho, Randy Cahya Wihandika, Dahnial Syauqy, Putra Pandu Adikara

Computer Vision Research Group, Faculty of Computer Science, Brawijaya University Malang, Indonesia

f3_ningrum@ub.ac.id

***Abstract*— Moving human is quite complicated to track since there are variations in background, texture, and lighting in an environment. This paper presents an effective method for tracking a human guide from a camera follower in both indoor and outdoor condition. This algorithm is designed to be embedded in a smart wheelchair. A conventional human detection by using Histogram of Oriented Gradient (HOG) was used at the first stage, then each detected human by HOG is utilized for tracking algorithm. The detected area from HOG is converted to grayscale image and its Entropy Difference Minimization (HOG-EDM) is calculated. The process is repeated for every frame. The entropy minimization is used as matching function in the tracking subsystem to determine the candidate of tracked object in the upcoming frame. The proposed algorithm has been proven to work well in indoor and outdoor area, even with textured background. Our testing based on self-made and public datasets shows that HOG-EDM method reaches over 80% accuracy.**

***Index Terms*— Camera Follower; Histogram Of Oriented Gradient; Human Tracking; Moving Camera.**

I. Introduction

People who cannot walk because of physical illness needs other people to help them in moving from one place to other. There are several attempt to develop equipment or software to bridge human with disabilities and computer. One of them is mini electric seater who could help complex disability sufferer to move with their eye movement [1]. The vehicle provides control by means of eye movement. Nowadays, the similar systems to mini electric seater has been developed in collaborative work by many researches. They build a prototype or algorithm embedded in a wheelchair to help the disabled person to do daily activities as common people [2][3]. In previous research, there was an attempt to guide a wheelchair using a line laser to estimate the distance between the wheelchair and obstacles [4]. The method was good for certain condition, such as indoor and outdoor in the afternoon. In order to extend previous research by adding its functionality, we decide to supplement the line laser with a front facing camera. This camera is utilized to detect and follow a defined guide in various lighting and background condition. Therefore, research related to human tracking is applied in this paper for upgrading the purpose of smart wheelchair to help disabled people to move flexibly. The system is aimed to lower the exhaustion level of a caregiver while looking after a person. For that reasons, a system equipped with human tracking sensor using camera follower is proposed.

Mobility has been providing support for people's quality of life [5][6]. Thus, it is essential for human to be able to move independently. Wheelchair is one of alternative equipment which often used to help mobility for such disabled patient. The conventional wheelchair can only be operated by those people who still can move their hand and fingers. Hence, people with hand and foot impairment cannot use the wheelchair by himself/herself. People with hand and foot impairment need an assistant to push the wheelchair, he/she cannot move freely. Based on the problem of our research, we implement our human tracking method in the smart wheelchair that can automatically following the movement people as an assistant at the position in front of a wheelchair. So, an assistant is not necessary to push the wheelchair and free to activity such as buy something at the supermarket, look around the tourism object and others activity. Hence, we offer a solution to support our smart wheelchair. The human tracking method is used to support in navigation system in the smart wheelchair.

Moving object detection is a part of basic issues in computer vision and video processing [7]. Moving human is quite difficult to track due to several problems in detecting or recognizing human in a video sequence. For instance, if we use moving camera, the background may be changed, or the environment condition may fluctuate rapidly [8]. Another example is overlapping human object may occur when a smart wheelchair follows a guide and could make the detection fail. In addition, some of researchers use point-based feature [9][10], in which every object is represented by a set of points as a feature descriptor. The problem of using key point feature is when a human does not have significant texture, for instance when they wear a plain motif of clothes. Another issue of using those kind of feature is high time complexity since each point has to be matched. Instead of using key point feature some researches use a sophisticated algorithm, which is HOG. HOG is utilized for content image authentication with gradient based feature extraction [11][12]. HOG could cope the problem as well for overlapping human object detection [13].

Apart from HOG, entropy represents the magnitude of each point location based on its likelihood. Shannon Entropy is one of methods that explains how much the information exists in the image [14]. Shannon entropy is applied for image analysis both in color or grayscale image as well. In addition, entropy is utilized for calculating the difference image for calibrating color image in unstable lighting condition which robust to be used as feature [15]. For static or moving camera a method that is robust from unstable lighting environment is a necessary part.

Based on the above reason, the authors proposed a combined robust method, HOG and EDM for tracking human guide by a camera follower in order to help disabled person who use wheelchair in their daily activities. We implement our human tracking method in the smart wheelchair that can automatically following the movement people as an assistant at the position in front of a wheelchair. So, an assistant is not necessary to push the wheelchair and free to activity such as buy something at the supermarket, look around the tourism object and others activity. Hence, we offer a solution to support our smart wheelchair. The human tracking method is used to support in navigation system in the smart wheelchair.

The rest of this paper is organized as follows. Section 2 gives the detail of dataset information and steps of proposed algorithm. Section 3 describes the result and discussion. Finally, the paper is concluded in Section 4.

II. METHOD OF RESEARCH

The dataset used in this research contains 5 video files, two of them are taken from public videos available on the Internet and the rest are self-made dataset recorded indoor without taking care about the lighting condition. The public dataset were taken from pixabay.com which have free license to use. Various conditions of both types of dataset is used to make sure that EDM algorithm is sufficiently robust for any situation. The detailed information of the dataset is shown at Table 1.

The self-made dataset was captured using a web camera which is embedded in the smart wheelchair with static and dynamic position. The experiment was taken place in Python environment paired with OpenCV library.

The general method is explained in the Figure 1. The input of the system is video where HOG is applied at the first phase for detecting human. It is possible to have more than one detected human in one frame, thus to decide a specific guide, the user must calibrate the system first. The user voluntarily selects which person to follow as a guide.

Based on the reference [16], upper body feature is more accurate to be employed in human tracking algorithm, because for some real cases, the lower body is inclined to be covered by other objects. The upper body feature is more useful for the systems that will be embedded in the smart wheelchair to maintain the distance between the guide and the wheelchair. If the camera follower is set to recognize the whole body of the guide, the distance between human and a smart wheelchair may be too far and it could be not ideal because of the disabled person in the wheelchair must be supervised by the guide.

After one third of upper body area is detected via HOG, the detected chunk is converted into grayscale. The grayscale conversion is intended to capture more homogeneous feature from a color image. Then, histogram from the grayscale image is built to compute entropy. Entropy is calculated for each detected object by HOG in every frame. The area for entropy calculation is limited to the one third upper part of HOG detection result to get constant result in each frame.

Table 1
Description of video dataset

State	Number of frames	Notes
Video 1: Indoor with moving camera (self-made dataset) 	785	Camera is dynamically following a woman to move from one room to go outside. The speed of movement is medium with a halt for a second because the woman met another man and spoke to each other, then continued to walk. There was a change in background, but not too significant. There was a bit overlapping objects.
Video 2: Indoor with static camera (self-made dataset) 	583	Camera was static which laid on smart wheelchair following a woman and a man who walked together a long way in the room. The speed of movement was quite fast and there was overlap a human or object who walked crossing behind a woman and a man. The background was not much textured. A woman wore a scarf and a man was in a shirt with no extreme motif.
Video 3: Indoor with static camera (self-made dataset) 	507	Camera was static laying on smart wheelchair following two men without any overlapping object. The speed of movement was medium. Two men wore a shirt with almost identical color. The background was static without any noticeable texture.
Video 4: Outdoor at morning or afternoon (public dataset taken from pixabay.com) 	859	Camera was static without any overlapping object. A woman walked to enter the scene and continuing walking away from camera and the speed of her movement was fast. The background tend to be static, except for several frames showing opening fences.
Video 5: Outdoor at night with an enough lighting (public dataset taken from https://pixabay.com/en/videos/walk-couples-in-love-love-3974/) 	2188	Camera was static capturing a couple walking together. The couple walked until far away from the camera and another couple walked together as well behind them, thus a bit overlapping is occurred. The speed of movement was fast. The background in each frame tend to be static without any extreme lighting.

Another problem of detecting or choosing guide is caused by the detection algorithm that solely use HOG method. If a frame could not detect a human in certain frame, the systems scans the next frame until a specific area of a human is detected. This condition always happened for the first time a guide is registered. After registering guide, the proposed algorithm will be executed by comparing the entropy feature in specific guide area.

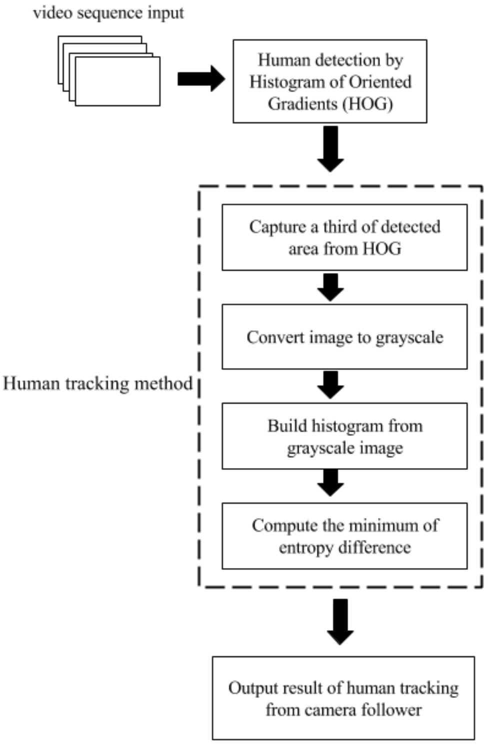


Figure 1: Proposed Method

A. Human Detection using HOG

HOG algorithm is adequately simple to understand. HOG is one of feature descriptor that is utilized in image processing and computer vision for recognizing or detecting object. In some cases, HOG is used as feature extraction method based on gradient of histogram for human detection. Then HOG is combined using Support Vector Machine (SVM) as the classification technique for features that already extracted. HOG feature descriptor is composed as a feature for similarity matching. Therefore, it could make the process of classifying more straightforward.

HOG use the global feature to recognize a human instead of using local feature. It creates a problem if a human using scarf or head cover. The identification of a human is wrong during unsuitable feature, however, because of HOG pays attention to the shape of the body, the algorithm still quite good to detect a human in particular case. The concepts of HOG are detecting feature in each sliding window of a picture, which is passed around image. The feature will be automatically classified by SVM into human or not human object. Figure 2 shows the example result of HOG. The algorithm of HOG as human detection is described as follows:

1) Separate the image area into a grid of points. For each point:
2) Partition a square window sized a x b in an image into a set of cells.
3) Calculate image gradient of a cell and apply a Gaussian window with sigma = 0.5 x a x b
4) Create a $a \times b \times b$ spatial and orientation histogram.
5) Apply tri-linear interpolation for each cell to build histogram using gradient magnitude.
6) Use L2-Hys normalization to each cell separately.
7) Merge HOGs from all cell into a descriptor vector.

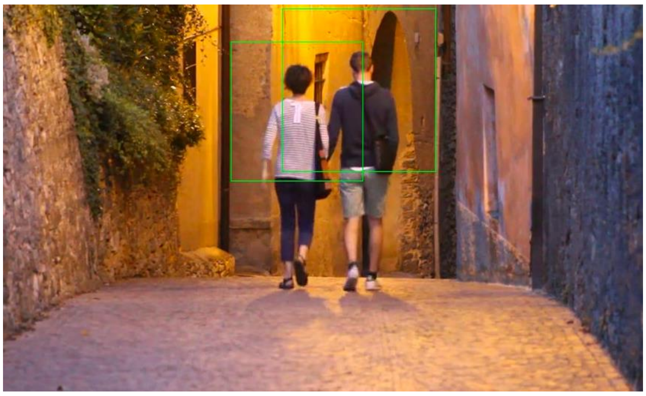


Figure 2: Example of the result of HOG

B. Human Tracking Method

Based on the result of human detection using HOG as shown in Figure 2, the proposed method was computed. The initialization at the first frame is needed to choose the specific guide that will be followed. However, the problem will come when at the first frame of the video, the HOG algorithm does not work. It is happened because of the recognizing of human is not match using HOG feature. Thus, the initialization of guide is decided based on frame which has the first detected human as guide. Equation (1) defines the process of finding initial guide using HOG.

$$F_f(x,y) = \begin{cases} F(x,y) & \textit{, human detected} \\ F_{f++}(x,y) & \textit{, find other frames} \end{cases} \quad (1)$$

where $F_f\,(x,y)$ is frame with successful detection. If the frame contains human, thus $F(x,y)$ defines as the first frame with detected human, otherwise the systems keep trying to find out human in the next frame. There are two possible conditions while finding a guide as follows:

a) If there is no detected area in the certain frame, thus we should move to the next frame to find a guide.
b) If there are two or more detected areas that represent human, we should choose one of them to be a guide.

After having a certain frame that contains guide area, this paper proposes a model of area that derives a specific feature that can be extracted using entropy. Figure 3 shows that we use a upper third part of detected block from HOG detection, which is extracted from a third of height and width, respectively. This approach is chosen because if we use the whole area of detected human, it generates noise, because the area contains a part of background as well. The background in the dataset has various degree of disturbance such as extreme light condition. Therefore, getting smaller areas in this case reduces noise that leads to higher accuracy in human tracking.

By taking a third of detected area, then we are required to determine the divergence measures of each detected area features to ones of the next frame. The equation (2) describes how the Shannon entropy is calculated. Shannon entropy is employed after converting color images to grayscale ones, due to the environment of dataset that have extreme color difference. Thus, by using grayscale image, the color feature of the image is homogenized.

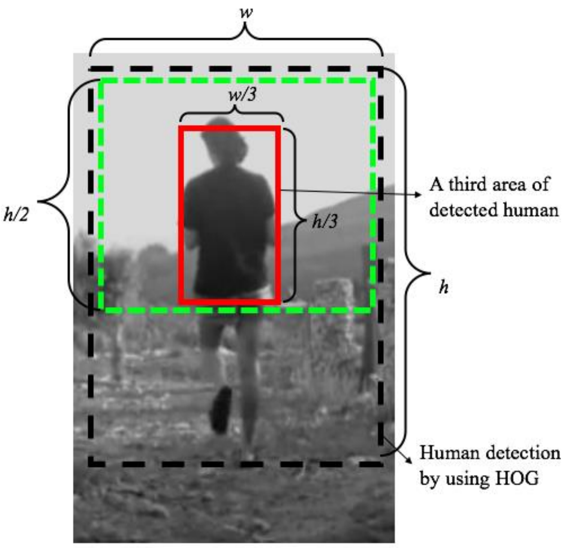


Figure 3: A third area of detected area

$$H_g = -\sum_{i=0}^{255} p_i log_2 \, p_i \tag{2}$$

where *Hg* is entropy from a third area of detected image in each frame of sequence video and *pi* is frequency normalization from the occurrence of intensities of grayscale image.

For summarizing the information among the intensities, the average distribution in each pixel per bits is calculated [14]. Entropy is sufficiently robust for feature extraction in grayscale image. The minimization difference among a third detected area in each frame is defined in order to track the guide. Equation (3) shows the calculation of the difference in the sequential frame in the video.

$$F_{gt}(x, y) = \Delta(H_{gt}(x, y))$$
$$F_{gt}(x, y) = (H_{gt}(x, y) - H_{gt-1}(x, y) \tag{3}$$

where $F_g(x,y)$ is temporal frame of a third area entropy at time *t* and $H_g(x,y)$ is the entropy of intensity pixel *(x,y)* in the frame at time *t* or current frame. On the other hand, time *t-1* is the previous frame.

III. Experimental Result and Discussion

To evaluate performance of the proposed method for tracking human guide under unstable lighting and background environment. The testing scenario are based on the number of frames which contains a guide detected by HOG and the number of frames which contains guide sequentially detected by HOG and EDM as a powerful method for detection and tracking.

There are 5 videos tested using the proposed algorithm in numerous condition as described in Table 1. In an entire videos, only several frames containing human and EDM always depends on whether there is human or not in a frame. In some cases, there are some deviation in the dataset such as overlapping human or more than one human are detected in a frame.

Table 2
The accuracy of human guide algorithm using self-made and public dataset

Video	Number of frames	Correctly detected frames		Accuracy of detected frames	
		HOG	HOG-EDM	HOG	HOG-EDM
Video 1	38	13	25	0.3421	0.6579
Video 2	45	27	37	0.6000	0.8222
Video 3	53	31	36	0.5849	0.6792
Video 4	145	124	137	0.8552	0.9448
Video 5	427	97	404	0.2272	0.9461
Average				**0.5219**	**0.8101**

Table 2 shows the accuracy for each video as well as the average accuracy for whole dataset. The result shows that 80% of accuracy was achieved by using combined HOG-EDM during the test. In the other hand, HOG alone only achieved 50% accuracy. It proves that HOG-EDM overcomes the problem of guide tracker under indoor and outdoor as well as invariant to lighting environment, over the conventional method as HOG. Average result concludes that HOG-EDM is applicable to be embedded in smart wheelchair as main component of guiding disabled with hands-free movement.

The specific or detail result is described in the Figure 4, 5, 6, 7, and 8 respectively. Figure 4 shows a human guide tracking for a woman who wearing scarf, the accuracy of detection reach around 66%. Human detection using HOG in this video achieve only 38 frames out of 785 frames. The reflection from the window cause improper recognition. The HOG feature could detect either a guide or other object such as window. When there is a man across the camera as shown in Figure 4 (e) the HOG has false detection. The detector starts to recognize the man for few frames. However, HOG still performs well because in the following frames, the human guide detection back to the woman. EDM run smoothly even when the woman moved fast at the first frame and halted for a second, then continue to walk in slow speed. HOG may detect one or more objects as human, but by combining with EDM the tracker could stick to follow only one guide in one frame.

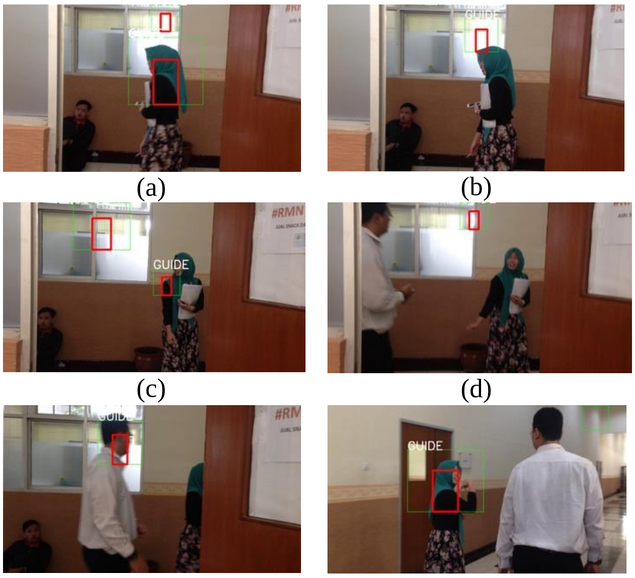
Figure 5 produces great accuracy over 80% and follows a human guide, even when there are three people the EDM is constantly good for tracking. Same as Video 1, there is people who quickly crossing the two people and automatically detected. There are only 8 frames having false detection and giving track to a guide and the rest is still robust. Based on the analysis, the drawback of tracking is because of similar color skin feature between a guide and other people. Skin and hair color of those people are similar, as well as their clothes which has similar plain motif.

Similar abnormalities also appear in Video 3, two men walking together with no significant difference in background texture. The static camera is used and the lighting condition is not bright enough, thus could make the accuracy less than 70%.

Figure 7 shows that the tracking was successful when detecting one human only in environment. There is no much noise of another object aside from human guide. There is a fence which often moves in the video. The environment seems has bright light with some of part having reflection or shadow. The detection and tracking yield over than 90% which constantly moving from nearby until go far away from static camera.

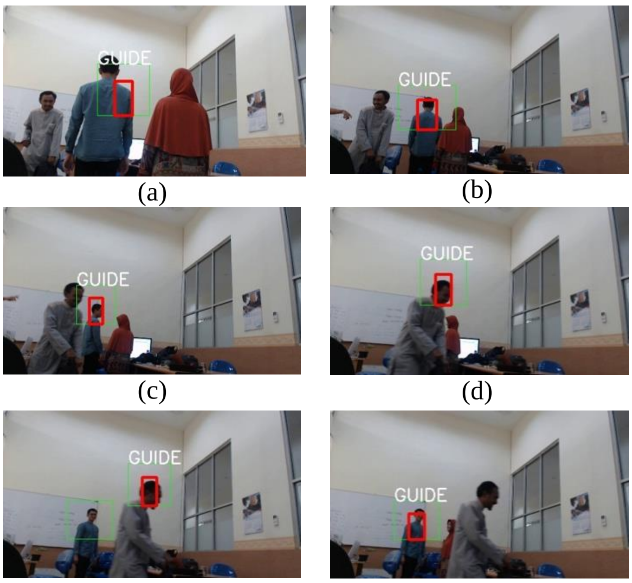
Still for outdoor area, the Video 5 shows two people walking together at night with lighting using lamp. The lighting is a bit yellowish, however the entropy was extracted very good, because of the women wear textured which easily to detect the unique guide comparing to the other features in a frame. The accuracy in this video could reach over than 90% which could follow a guide from nearby until goes long away from static camera. When another couple walking behind a guide, the detection is still robust to keep recognizing a guide.

Although HOG-EDM can be applied in the different lighting condition, however this method remain poses a drawback when there are two or more people are overlapping during the detection and recognition. The indexing of detected box is arbitrarily changing because of the value of the entropy feature of current frame and the previous one is only slightly different. Thus, the improvement is needed to reveal more discriminative features for the detection.



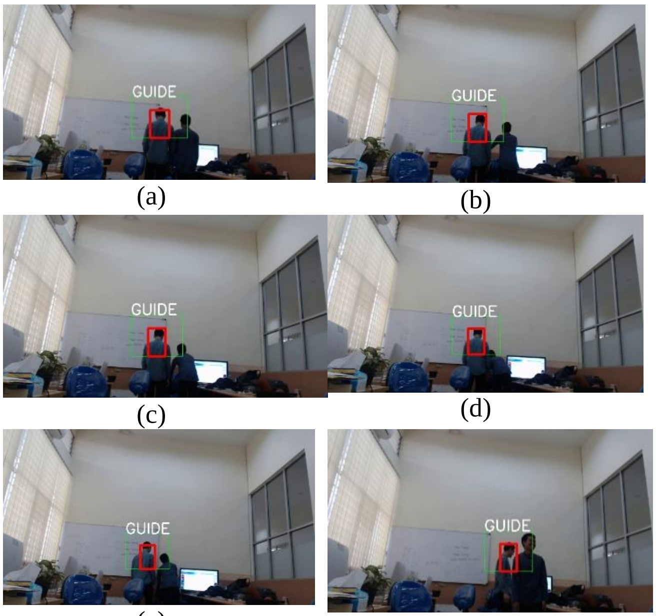
(a) (b) (c) (d) (e) (f)

Figure 4: Human detection and tracking in Video 1



(a) (b) (c) (d) (e) (f)

Figure 5: Human detection and tracking in Video 2



(a) (b) (c) (d) (e) (f)

Figure 6: Human detection and tracking in Video 3

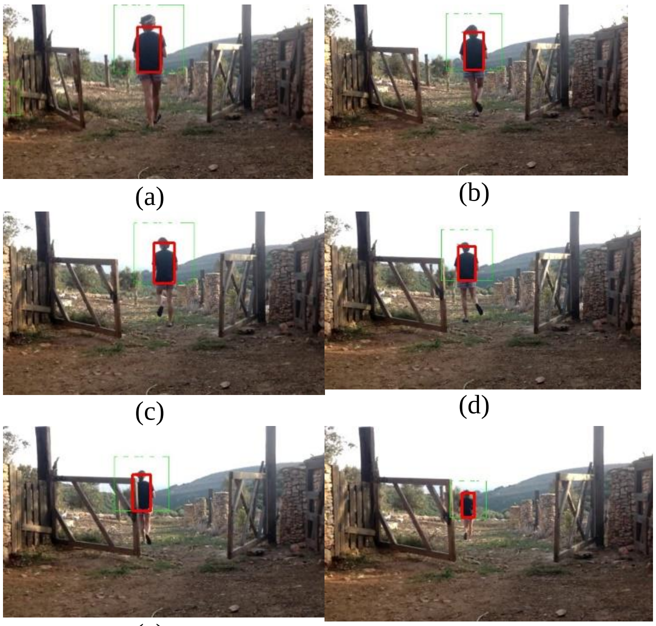

(a) (b) (c) (d) (e) (f)

Figure 7: Human detection and tracking in Video 4

IV. CONCLUSION

In this paper, a novel method for human guide tracking is proposed by means of combined HOG and EDM. Entropy measure is a feature that could interpret the information of specific area of image through from its probability of intensities in grayscale image. This algorithm has drawback when there are overlapping human object. Some of frames detect wrong guide or detect more than one guide, even more for people who has same or similar skin color. The feature is still not covered those kinds of issues.

In the other hand, HOG-EDM could be used in different lighting condition, both for indoor and outdoor. It gains very good result if the human guide wearing textured clothes, because it is easier for entropy to keep the information rather than wearing plain color. From whole experiments, HOG-

EDM obtained over 80% for accuracy. It shows that HOG-EDM is sufficient robust for any cases with unstable lighting condition and a bit movement of camera follower. This algorithm also stable for guide with no extreme background and achieved very good if a human guide is very unique feature than others human detected.

The future work will be applied additional specific feature in order to get more robust tracking, moreover for feature with plain clothes and a guide with scarf or head covered. The HOG algorithm could be optimized as well for finding a right guide and track a guide constantly along the frames of the video.

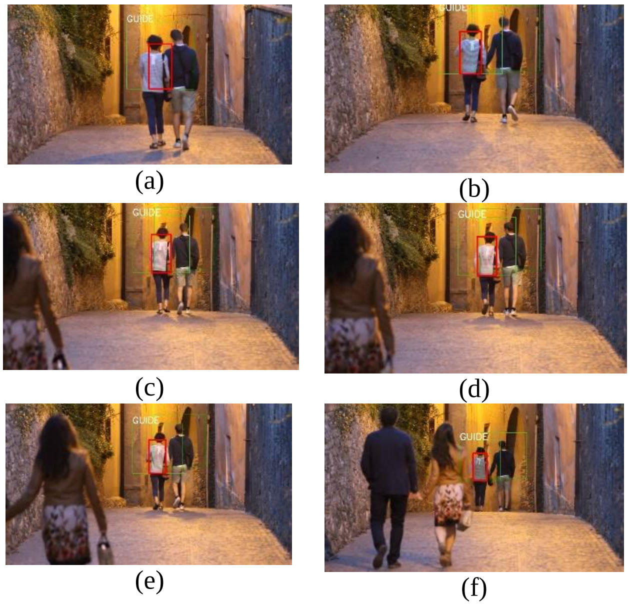


Figure 8: Human detection and tracking in Video 5

ACKNOWLEDGMENT

We would like to thank to all of members in Computer Vision Research Group, Faculty of Computer Science, Brawijaya University Indonesia who dedicated to help author for preparing the dataset as well as the Minister of Higher Education Indonesia for supporting the research fund.

REFERENCES

[1] Utaminingrum, Fitri, M. Ali Fauzi, Yuita Arum Sari, Renaldi Primaswara, and Sigit Adinugroho. "Eye Movement as Navigator for Disabled Person." In Proceedings of the 2016 International Conference on Communication and Information Systems, pp. 1-5. ACM, 2016.

[2] Leaman, J., La, H. M., & Nguyen, L. (2016, September). Development of a smart wheelchair for people with disabilities. In Multisensor Fusion and Integration for Intelligent Systems (MFI), 2016 IEEE International Conference on (pp. 279-284). IEEE.

[3] Dalsaniya, A. K., & Gawali, D. H. (2016, January). Smart phone based wheelchair navigation and home automation for disabled. In Intelligent Systems and Control (ISCO), 2016 10th International Conference on (pp. 1-5). IEEE.

[4] Utaminingrum, Fitri, Tri Astoto Kurniawan, M. Ali Fauzi, Rizal Maulana, Dahnial Syauqy, Randy Cahya Wihandika, Yuita Arum Sari, Putra Pandu Adikara. "A Laser Vision based Obstcale Deection and Distance Estimation for Smart Wheelchair Navigation". In 2016 IEEE International Conference on Signal and Image Processing, pp. 123-127.

[5] Utaminingrum, M. Ali Fauzi, Randy Cahya Wihandika, Sigit Adinugroho, Dahnial Syauqy, Yuita Arum Sari, Putra Pandu Adikara, "Development of Computer Vision Based Obstacle Detection and Human Tracking on Smart Wheelchair for Disabled Patient". International Symposium on Computational and Business Intelligence (ISCBI) 2017.IEEE

[6] Utaminingrum, M. Ali Fauzi, Randy Cahya Wihandika, Putra Pandu Adikara, 2016, "Adaptive Human Tracking for Smart Wheelchair". International Symposium on Computational and Business Intelligence (ISCBI), Aug 2017.IEEE

[7] Yanmin, L., Shujiao, J., Xiuli, G., Zhibin, F., Xiaoxue, X., & Limin, D. (2016, May). Research on automatic tracking caregivers of intelligent wheelchair. In Control and Decision Conference (CCDC), 2016 Chinese (pp. 801-806). IEEE.

[8] Hu, W. C., Chen, C. H., Chen, T. Y., Huang, D. Y., & Wu, Z. C. (2015). Moving object detection and tracking from video captured by moving camera. Journal of Visual Communication and Image Representation, 30, 164-180.

[9] Sincan, O. M., Ajabshir, V. B., Keles, H. Y., & Tosun, S. (2015, September). Moving object detection by a mounted moving camera. In EUROCON 2015-International Conference on Computer as a Tool (EUROCON), IEEE (pp. 1-6). IEEE.

[10] Wang, H., & Yi, Y. (2015, October). Tracking Salient Keypoints for Human Action Recognition. In Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on (pp. 3048-3053). IEEE.

[11] Shin, J., Kim, D., & Ruland, C. (2014, October). Content based image authentication using HOG feature descriptor. In Image Processing (ICIP), 2014 IEEE International Conference on (pp. 5292-5296). IEEE.

[12] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.

[13] Ghorbani, M., Targhi, A. T., & Dehshibi, M. M. (2015, October). HOG and LBP: Towards a robust face recognition system. In Digital Information Management (ICDIM), 2015 Tenth International Conference on (pp. 138-141). IEEE.

[14] Kwok, N., Shi, H., Ha, Q., Fang, G., Chen, S. and Jia, X. (2013). "Simultaneous image color correction and enhancement using particle swarm optimization". Engineering Applications of Artificial Intelligence, 26 (10), pp. 2356-2371.

[15] Sari, Y. A., Ginardi, R. H., & Suciati, N. (2015, September). Color correction using improved linear regression algorithm. In Information & Communication Technology and Systems (ICTS), 2015 International Conference on (pp. 73-78). IEEE.

[16] Li, C., Shi, X., Xu, Y., & Wu, S. (2012, October). Human body detection and tracking from moving cameras. In Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on (pp. 278-281). IEEE.