Memory-Efficient Implementation of GMM-MRCoHOG for Human Recognition Hardware

Ryogo Takemoto[1] [a], Yuya Nagamine[1], Kazuki Yoshihiro[1], Masatoshi Shibata[2], Hideo Yamada[2],
Yuichiro Tanaka[3] [b], Shuichi Enokida[4] [c] and Hakaru Tamukoh[1,3] [d]

[1]*Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,*
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan
[2]*AISIN CORPORATION, 2-1 Asahi-machi, Kariya, Aichi, 448-8650, Japan*
[3]*Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology,*
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan
[4]*Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology,*
680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan

Keywords: Image Processing, Human Recognition, Human Detection, HOG, MRCoHOG, GMM-MRCoHOG, FPGA.

Abstract: High-speed and accurate human recognition is necessary to realize safe autonomous mobile robots. Recently, human recognition methods based on deep learning have been studied extensively. However, these methods consume large amounts of power. Therefore, this study focuses on the Gaussian mixture model of multiresolution co-occurrence histograms of oriented gradients (GMM-MRCoHOG), which is a feature extraction method for human recognition that entails lower computational costs compared to deep learning-based methods, and aims to implement its hardware for high-speed, high-accuracy, and low-power human recognition. A digital hardware implementation method of GMM-MRCoHOG has been proposed. However, the method requires numerous look-up tables (LUTs) to store state spaces of GMM-MRCoHOG, thereby impeding the realization of human recognition systems. This study proposes a LUT reduction method to overcome this drawback by standardizing basis function arrangements of Gaussian mixture distributions in GMM-MRCoHOG. Experimental results show that the proposed method is as accurate as the previous method, and the memory required for state spaces consuming LUTs can be reduced to 1/504th of that required in the previous method.

1 INTRODUCTION

The demand for home service robots (Iocchi et al., 2015) (Yamamoto et al., 2019) (Ono et al., 2022) (Yoshimoto and Tamukoh, 2021) and self-driving cars (Gupta et al., 2018) (Fei et al., 2021) (Bojarski et al., 2016) has been increasing owing to the accelerating aging of society and the declining birthrate. High-speed and high-accuracy human recognition processing is required for realizing safe autonomous mobile robots that can coexist with humans. Currently, most human recognition implementation systems are based on deep learning (Hinton et al., 2006) (Lecun et al., 1998) (Krizhevsky et al., 2012) (Redmon et al., 2016) (Redmon and Farhadi, 2018) (Bochkovskiy et al., 2020) using graphics processing units (GPUs), which can achieve state-of-the-art accuracy with real-time processing. However, they have the disadvantages of high power consumption and considerable heat generation, making it difficult to implement these systems in robots. A dedicated hardware implementation is one of the solutions for these problems. Furthermore, a low computational cost algorithm, unlike deep learning-based methods that incur high computational costs, is desirable because of limited hardware resources in mobile robot systems.

Several studies have adopted hardware implementation for high-speed and low-power robot systems. For example, Ishida *et al.* proposed hardware in-
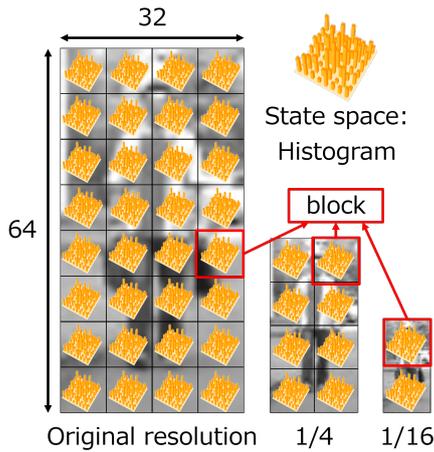
[a] https://orcid.org/0000-0002-6795-0794
[b] https://orcid.org/0000-0001-6974-070X
[c] https://orcid.org/0000-0001-6309-3185
[d] https://orcid.org/0000-0002-3669-1371

Figure 1: Luminance gradient co-occurrence histograms for state spaces in MRCoHOG.

Figure 2: Gaussian mixture distributions for state spaces in GMM-MRCoHOG.

telligent processing accelerator based on a field programmable gate arrays (FPGA) (Ishida et al., 2020). Tanaka *et al.* proposed a brain-inspired artificial intelligence model based on FPGAs for home service robots (Tanaka et al., 2020). Both studies proposed hardware-oriented algorithms that reduced computational costs.

For human recognition, Takemoto *et al.* (Takemoto et al., 2022) utilized a hardware implementation of the Gaussian mixture model-multiresolution co-occurrence histograms of oriented gradients (GMM-MRCoHOG) algorithm (Higashi et al., 2018) (Nagamine et al., 2021), which performs human recognition with lower computational costs compared to deep learning-based methods. GMM-MRCoHOG is a derivative algorithm of MRCoHOG (Iwata and Enokida, 2014) that extracts useful features for human recognition by accumulating luminance gradient co-occurrence into a histogram in each block of an image, called a state space, as shown in Figure 1. GMM-MRCoHOG optimizes the state space by approximating the histogram with a Gaussian mixture distribution, as shown in Figure 2. It uses less memory and is more accurate than MRCoHOG.

Takemoto *et al.* proposed a hardware-oriented algorithm for embedded systems that reduces the hardware resources required by simplifying complex operations in the original GMM-MRCoHOG algorithm. However, this method still requires many look-up tables (LUTs) to store state spaces. The state space shape differs for each block as each state space is constructed with a Gaussian mixture distribution; therefore, many LUTs may be required based on the total number of blocks. LUT reduction is necessary to achieve scalable human recognition systems because the number of LUTs required increases as the image size increases.
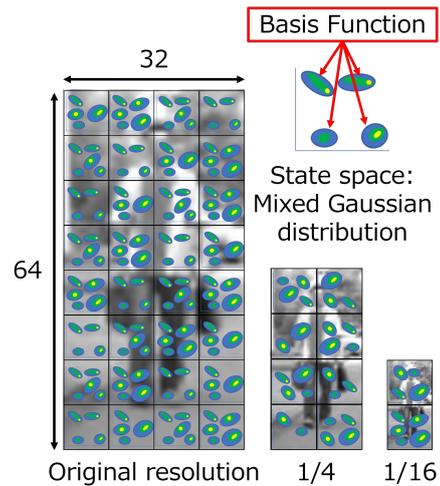
Therefore, we propose an LUT reduction method wherein basis function arrangements of Gaussian mixture distributions of the GMM-MRCoHOG state spaces are standardized. The number of LUTs can be reduced by aggregating the features of all blocks into a single space and constructing a single Gaussian mixture distribution for all state spaces.

2 RELATED WORKS

2.1 MRCoHOG

MRCoHOG, a derivative algorithm of HOG (Dalal and Triggs, 2005), downsamples images in two steps, as shown in Figure 1, and extracts features by representing the luminance gradient co-occurrence among images of three resolutions as a two-dimensional histogram. For example, if the dataset image size is 32×64 pixels, the number of blocks at each resolution is $8 \times 4 = 32$, $4 \times 2 = 8$, and $2 \times 1 = 2$.

For reducing computational cost, pairs of co-occurrence gradients, called offsets, are limited to 36 pairs of a pixel of interest and its four neighboring pixels among three resolutions, as shown in Figure 3. However, methods that use histograms as features, such as HOG, require designers to discretize the luminance gradients; the discretization error of gradient information and feature generalization ability depend on the class width. Manually determining the optimal class width is difficult.
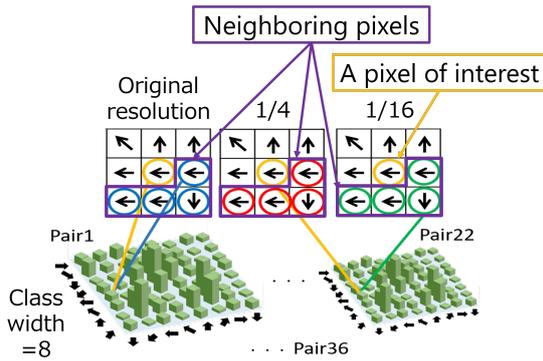
Figure 3: Luminance gradient co-occurrence accumulation of the MRCoHOG algorithm.

2.2 GMM-MRCoHOG

To solve the problem of optimal class width determination in MRCoHOG, GMM-MRCoHOG autonomously constructs luminance gradient co-occurrence histograms as state spaces by approximating them using Gaussian mixture distributions and then extracts input image features based on the distributions.

Figures 4 (a) and (b) show a state space decision process using positive and negative data in the training phase of GMM-MRCoHOG. Luminance gradients of a training image are discretized in 36 directions, and the gradient co-occurrence is plotted in state spaces for the positive and negative data. Then, the positive and negative data distributions are approximated via Gaussian mixture distributions. Next, the Jensen-Shannon (JS) divergence, a measure of the difference between two probabilities of occurrence, is used to generate a new Gaussian mixture distribution that separates the positive and negative data distributions in a single space (Michishita et al., 2018). An absolute value of the JS divergence increases as the shapes of the two Gaussian mixture distributions differ. Some data distributions that strongly characterize both positive and negative data are obtained by extracting areas with high absolute values from the two Gaussian mixture distributions. For this, the inversion method, a random number generation method, is used to generate samples based on the JS divergence of the positive and negative Gaussian mixture distributions, resulting in several samples tending to be in areas with a strong bias toward either positive or negative data. Then, the EM algorithm (Dempster et al., 1977) is used to approximate the distribution to a Gaussian mixture distribution.

The generated Gaussian mixture distribution is used for feature extraction. Figure 5 illustrates the process of feature extraction from an input image dur-
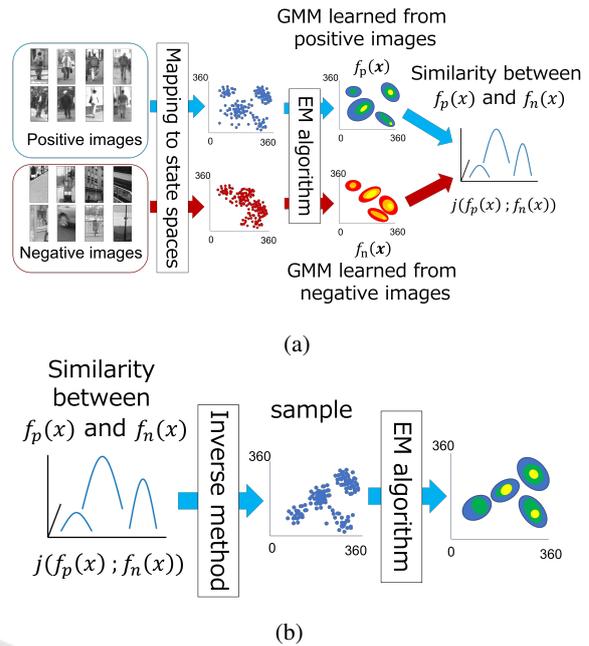

(a)

(b)

Figure 4: State space decision process in GMM-MRCoHOG.

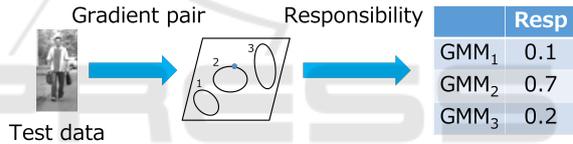|	Resp
GMM_1	0.1
GMM_2	0.7
GMM_3	0.2

Figure 5: Feature extraction process in GMM-MRCoHOG.

ing the inference phase of GMM-MRCoHOG. The feature value of an input luminance gradient pair is represented as responsibility of basis functions of the Gaussian mixture distribution. In GMM-MRCoHOG, the number of feature dimensions is determined by the number of basis functions of the Gaussian mixture distribution and does not depend on the class width. The number of basis functions is also called the number of Gaussian mixture distribution mixtures.

The number of mixtures of Gaussian distributions differs among blocks because the algorithm independently optimizes a state space for each block. In MR-CoHOG, memories to store 64 bins are allocated for every block when using 8×8 histograms. Conversely, some state spaces may have a small number of basis functions when an optimal state space is configured for each block, as in GMM-MRCoHOG. Thus, memory utilization for state spaces can be reduced. Furthermore, Gaussian mixture distributions also allow enable more precise representation of features than histograms, and GMM-MRCoHOG is more accurate than MRCoHOG.

2.3 Hardware-Oriented GMM-MRCoHOG

A hardware-oriented algorithm that simplifies complex operations in the original algorithm is necessary for the high-speed and low-power hardware implementation of GMM-MRCoHOG. Takemoto *et al.* proposed a hardware-oriented GMM-MRCoHOG (Takemoto et al., 2022), which is a simplified version of GMM-MRCoHOG, for FPGA implementation. In the algorithm, the luminance gradient computation in 36 directions includes complex operations, such as divisions and nonlinear function calculations, and is simplified using a coarse angle computation method based on a fixed-point $\tan\theta$ comparison table. Additionally, they designed a human recognition architecture using the simplified algorithm and pipeline processing.

This section describes the coarse angle calculation method in detail. An original angle calculation in GMM-MRCoHOG includes $\tan^{-1}\theta$ to compute a luminance gradient angle θ from horizontal and vertical luminance gradients f_x and f_y, respectively. Conversely, in the coarse angle calculation method, assuming that angle θ appears in the first quadrant, discretized $\tan\theta(\theta = 0, 10, ..., 80)$ are calculated in advance, and a $\tan\theta$ comparison table is constructed based on the relationship between the luminance gradient f_y/f_x and the discretized $\tan\theta$, as in Eq. (1). The second to fourth quadrants can be calculated similarly using the symmetry of trigonometric functions.

$$
\begin{aligned}
&if\ \tan 0° \leq \frac{f_y}{f_x} < \tan 10° \\
&\quad direction = 1(\theta : 0° \sim 10°) \\
&elif\ \tan 10° \leq \frac{f_y}{f_x} < \tan 20° \\
&\quad direction = 2(\theta : 10° \sim 20°) \\
&\qquad \vdots \\
&elif\ \tan 80° \leq \frac{f_y}{f_x} \\
&\quad direction = 9(\theta : 80° \sim 90°)
\end{aligned}
\tag{1}
$$

Additionally, the division in Eq. (1) is eliminated for simplifying the algorithm to reduce the hardware resources required. The $\tan\theta$ comparison table shown in Eq. (1) can be replaced with that shown in Eq. (2) because $f_x > 0$ and $f_y \geq 0$.

$$
\begin{aligned}
&if\ f_x \times \tan 0° \leq f_y < f_x \times \tan 10° \\
&\quad direction = 1(\theta : 0° \sim 10°) \\
&elif\ f_x \times \tan 10° \leq f_y < f_x \times \tan 20° \\
&\quad direction = 2(\theta : 10° \sim 20°) \\
&\qquad \vdots \\
&elif\ f_x \times \tan 80° \leq f_y \\
&\quad direction = 9(\theta : 80° \sim 90°)
\end{aligned}
\tag{2}
$$

Finally, the $\tan\theta$ comparison table is approximated using the fixed-point numbers, and the multiplication in Eq. (2) is replaced with a combination of bit-shift and additional operations, which requires fewer resources than floating-point multiplications.

However, this method still requires many LUTs to store state spaces because the state space shapes differ among blocks owing to the characteristics of the Gaussian mixture distribution. Specifically, the number of state spaces is 504 in the case of the same parameter setting as that in Section 2.1; the numbers of blocks in the images of three resolutions are 32, 8, and 2, and the number of offsets is 12. Takemoto et al. synthesized the hardware-oriented GMM-MRCoHOG using Vivado HLS 2018.2, and the circuit consumed 27,331 LUTs, which is more than 50% of LUTs available in an XC7Z020 FPGA on Xilinx ZedBoard (53,200 LUTs are available), limiting the implementation of multiple systems in an FPGA. Moreover, larger images are expected to be input owing to the recent camera performance improvements despite the input image size used in the study being 32×64 pixels, resulting in an increased number of blocks. Therefore, a LUT reduction method for the hardware-oriented GMM-MRCoHOG is required for system scalability.

3 PROPOSED METHODS

As a first step to implement GMM-MRCoHOG in hardware, this study proposes a novel hardware-oriented algorithm: a standardization method of basis function arrangements of Gaussian mixture distributions for all state space representations to reduce the number of LUTs required, which limits the previous hardware-oriented GMM-MRCoHOG. This method forms a single Gaussian mixture distribution by aggregating features of all blocks such that the distribution represents the state spaces of all blocks.

Similar to the original GMM-MRCoHOG, the proposed method autonomously constructs a state space using the EM algorithm, as shown in Figure 6. First, luminance gradient pairs of the positive and negative data of training images are extracted for each

block, and the corresponding positive and negative samples are approximated using Gaussian mixture distributions. Next, the JS divergence is used to generate a new Gaussian mixture distribution that separates the positive and negative data distributions in a single state space. To this end, the inversion method is used to generate samples based on the JS divergence of the positive and negative Gaussian mixture distributions, resulting in numerous samples tending to be present in areas with a strong bias toward either positive or negative data. Up to this point, each block is processed as in the original method. Next, the samples generated in all blocks in the previous procedure are aggregated into a single space to standardize the state spaces of all blocks, and a new Gaussian mixture distribution with the characteristics of all blocks is constructed autonomously using the EM algorithm. This single Gaussian mixture distribution is then used to represent the state spaces of all blocks.

The proposed method drastically reduces the memory requirement for state space representation. For example, in the previous method, if the maximum number of Gaussian mixture distribution mixtures is set to 8, the memory required for the state space is approximately 12 KB because the number of state spaces is 504, four variables representing each Gaussian distribution are required, and the bit width of each variable is 6 bits ($8 \times 504 \times 4 \times 6$ bits $\approx$ 12 KB). In contrast, the proposed method uses a common state space such that the memory utilization is 24 bytes ($8 \times 4 \times 6$ bits $\approx$ 24 B). Moreover, the modification of the proposed method does not affect the latency of the inference process because except for memory reading, it is the same as that of the previous method.

4 EXPERIMENT

We implemented the proposed standardization method of the basis function arrangements of Gaussian mixture distributions for all state spaces and conducted human recognition tasks. The experimental environment is presented in Table 1.

We evaluated the human recognition performance of the GMM-MRCoHOG using the proposed method. In this experiment, we compared the performance of the proposed method with that of the method pro-

Table 1: Experimental environment.

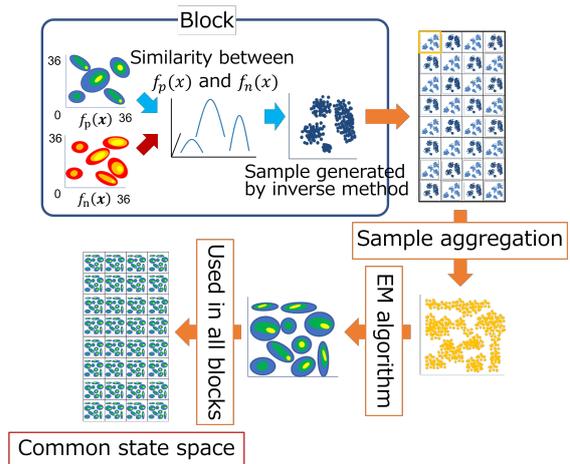| CPU	Intel Core i7-8700K 3.70 GHz
Memory	64 GB
Operating system	Windows 10
MATLAB version	R2021a

Figure 6: Proposed state space decision process through the standardization of basis function arrangements of Gaussian mixture distributions.

Figure 7: Examples of the INRIA Person dataset images.

Figure 8: Examples of the Daimler Pedestrian Classification Benchmark dataset images.

posed by Takemoto et al. The course angle computation described in Section 2.3 was applied to both methods. The maximum number of Gaussian mixture distribution mixtures was set to 8 or 16 for both the previous and proposed methods. The training data were obtained from the Daimler Pedestrian Classification Benchmark and INRIA Person datasets, and the testing data were obtained from the INRIA Person dataset. These datasets consist of human and non-human images of 32×64 pixels. Examples of images from the datasets are shown in Figures 7 and 8. A support vector machine (Cortes and Vapnik, 1995) with a linear kernel was used as the discriminator.

5 RESULTS

Figure 9 compares the performances of the previous and proposed methods, which are represented by receiver operator acting characteristic (ROC) curve. The vertical axis of the ROC curve indicates the true
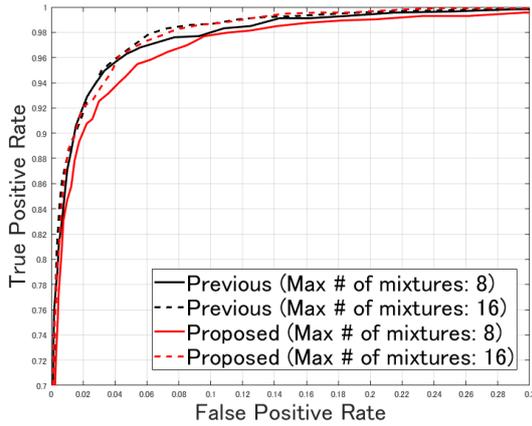
Figure 9: Comparison of the previous and proposed method performances.

Table 2: Human recognition task accuracies of the previous and proposed methods.

Maximum number of mixtures	Previous	Proposed
8	0.9668	0.9633
16	0.9685	0.9690

positive rate, and the horizontal axis indicates the false positive rate; the closer the curve is to the upper left, the higher the discrimination accuracy. The black and red lines indicate the performances of the previous and current methods, respectively. The solid lines indicate the method with the maximum number of mixtures set to 8, and the dashed lines indicate the method with the maximum number of mixtures set to 16. Table 2 presents the human recognition task accuracies of the previous and proposed methods. Figure 9 and Table 2 indicate that the proposed method was slightly inferior to the previous method when the maximum number of mixtures was 8 but had the same accuracy as the conventional method when the maximum number of mixtures was 16.

6 DISCUSSION

6.1 State Space Comparison

Figures 10 and 11 show the Gaussian mixture distributions of the first offset with the maximum number of mixtures set to 8 and 16 for the previous method, respectively. Figures 12 and 13 show the Gaussian mixture distributions with the maximum number of mixtures set to 8 and 16 for the proposed method, respectively. The experimental results indicate that the number of basis functions in the Gaussian mixture distributions was the same as the maximum number

of mixtures in all cases. Note that several basis functions are not displayed in these figures because they are overlapped or too small.

A comparison of these figures shows common characteristics in the arrangement of basis functions of the Gaussian mixture distributions in the state spaces. The basis functions tend to be concentrated on the diagonals and at the edges of the state spaces, even though the shapes of the state spaces are different, as shown in Figures 10 and 11. Figures 12 and 13 also show that the basis functions are placed on the diagonals and at the edges of the state spaces, as in the previous method. This indicates that the proposed method is as accurate as the previous method when the maximum mixing number is set to 16 because the Gaussian mixture distribution using the basis function arrangement can adequately represent the features of all blocks. However, the maximum number of mixtures of 8 is insufficient in terms of dimensionality to represent the features of all blocks. Similar tendencies were also observed for the other offsets.

6.2 LUT Utilization

In this experiment, the maximum number of mixtures of Gaussian mixture distributions in the state space for both the previous and proposed methods was set to 8 or 16. Experimental results showed that the number of basis functions in the Gaussian mixture distributions was the same as the maximum number of mixtures in all cases, implying that the number of LUTs used to store one state space was constant regardless of the method employed. Therefore, the number of LUTs for storing the state space in the proposed method was 504 times smaller than that in the previous method. Note that the human recognition archi-
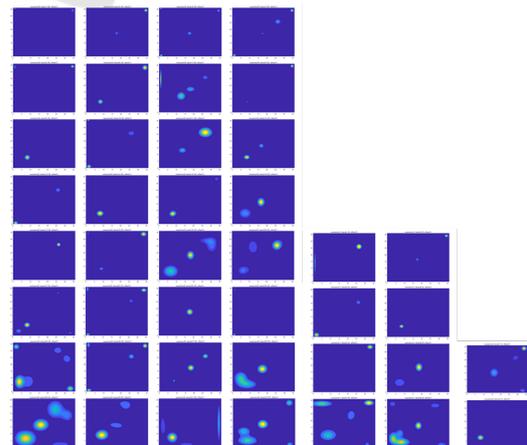

Figure 10: Basis function arrangements of the Gaussian mixture distributions in the previous method (maximum number of mixtures: 8, offset: 1).
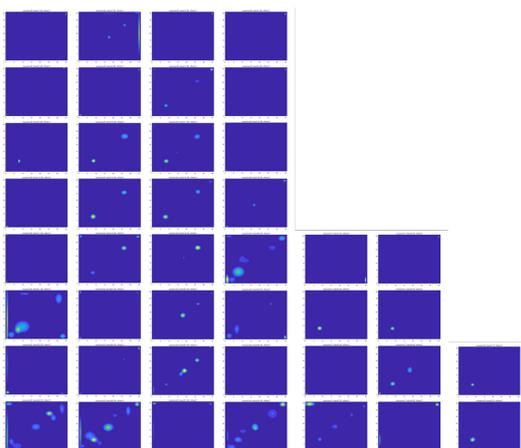
Figure 11: Basis function arrangements of the Gaussian mixture distributions in the previous method (maximum number of mixtures: 16, offset: 1).
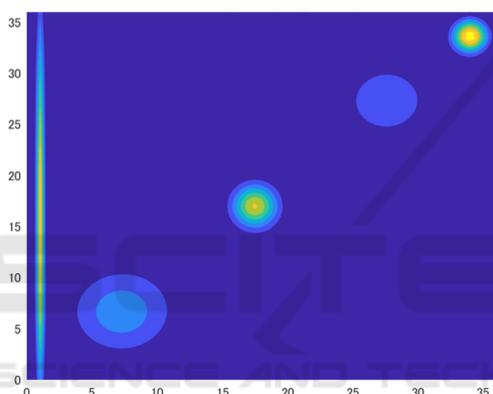

Figure 12: Basis function arrangements of the Gaussian mixture distributions in the proposed method (maximum number of mixtures: 8).

tecture has not yet been designed using the proposed algorithm, and the specific number of LUTs required for the entire system is unknown.

7 CONCLUSIONS

Human recognition with high speed, high accuracy, and low-power consumption is necessary to realize safe autonomous mobile robots that can coexist with humans. This study focused on GMM-MRCoHOG, which is capable of high-speed and high-accuracy human recognition, and aims to implement dedicated hardware to reduce power consumption. We proposed a standardization method of basis function arrangements of Gaussian mixture distributions that constructed the state space and confirmed that the number of LUTs required for the system was expected to be reduced.
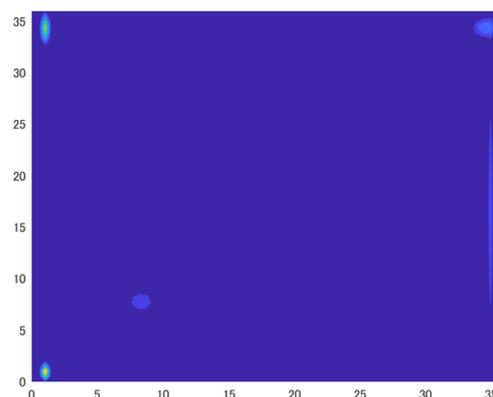

Figure 13: Basis function arrangements of the Gaussian mixture distributions in the proposed method (maximum number of mixtures: 16).

In the future, we will further verify the proposed hardware-oriented algorithm that can reduce the memory utilization for the state space representation and implement a human recognition architecture using the proposed hardware-oriented algorithm on an FPGA. We will continue to improve the proposed method by investigating the appropriate number of basis functions in the state space and examining the effect of changing the dataset. The human recognition architecture constructed using the proposed method is expected to require low memory and have high scalability. We will investigate the effectiveness of the system in the real world by mounting the FPGA on a robot.

REFERENCES

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J., et al. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Fei, J., Peng, K., Heidenreich, P., Bieder, F., and Stiller, C. (2021). PillarSegNet: Pillar-based semantic grid map estimation using sparse LiDAR data. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pages 838–844.

Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., and Alahi, A. (2018). Social GAN: Socially acceptable trajectories with generative adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2255–2264. IEEE Computer Society.

Higashi, S., Michishita, Y., Enokida, S., Shibata, M., and Yamada, H. (2018). Pedestrian detection based on Gaussian mixture model multiresolution CoHOG. In *Proceedings of the 4th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS)*, number MVML 100.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554.

Iocchi, L., Holz, D., del Solar, J. R., Sugiura, K., and van der Zant, T. (2015). RoboCup@Home: Analysis and results of evolving competitions for domestic and service robots. *Artificial Intelligence*, 229:258–281.

Ishida, Y., Morie, T., and Tamukoh, H. (2020). A hardware intelligent processing accelerator for domestic service robots. *Advanced Robotics*, 34(14):947–957.

Iwata, S. and Enokida, S. (2014). Object detection based on multiresolution CoHOG. In Bebis, G., Boyle, R., Parvin, B., Koracin, D., McMahan, R., Jerald, J., Zhang, H., Drucker, S. M., Kambhamettu, C., El Choubassi, M., Deng, Z., and Carlson, M., editors, *Advances in Visual Computing*, pages 427–437. Springer International Publishing.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Pereira, F., Burges, C., Bottou, L., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Michishita, Y., Higashi, S., Shibata, M., Muramatsu, R., Yamada, H., and Enokida, S. (2018). Autonomous state space construction method based on mixed normal distributions for pedestrian detection. In *IEEJ Transactions on Electronics, Information and Systems*, volume 138, pages 1100–1107.

Nagamine, Y., Yoshihiro, K., Shibata, M., Yamada, H., Enokida, S., and Tamukoh, H. (2021). A hardware-oriented algorithm of GMM-MRCoHOG for high-performance human detection by an FPGA. In Nakajima, M., Kim, J.-G., Lie, W.-N., and Kemao, Q., editors, *International Workshop on Advanced Imaging Technology (IWAIT) 2021*, volume 11766, page 117660B. International Society for Optics and Photonics, SPIE.

Ono, T., Kanaoka, D., Shiba, T., Tokuno, S., Yano, Y., Mizutani, A., Matsumoto, I., Amano, H., and Tamukoh, H. (2022). Solution of world robot challenge 2020 partner robot challenge (real space). *Advanced Robotics*, 36(17-18):870–889.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788. IEEE Computer Society.

Redmon, J. and Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Takemoto, R., Nagamine, Y., Yoshihiro, K., Shibata, M., Yamada, H., Tanaka, Y., Enokida, S., and Tamukoh, H. (2022). Hardware-oriented algorithm for human detection using GMM-MRCoHOG features. In Farinella, G. M., Radeva, P., and Bouatouch, K., editors, *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022*, volume 4: VISAPP, pages 749–757. SCITEPRESS.

Tanaka, Y., Morie, T., and Tamukoh, H. (2020). An Amygdala-inspired classical conditioning model implemented on an FPGA for home service robots. *IEEE Access*, 8:212066–212078.

Yamamoto, T., Terada, K., Ochiai, A., Sato, F., Asahara, Y., and Murase, K. (2019). Development of human support robot as the research platform of a domestic mobile manipulator. *ROBOMECH Journal*, 6(4).

Yoshimoto, Y. and Tamukoh, H. (2021). FPGA implementation of a binarized dual stream convolutional neural network for service robots. *Journal of Robotics and Mechatronics*, 33(2):386–399.