IEICE TRANSACTIONS

on Fundamentals of Electronics, Communications and Computer Sciences

VOL. E100-A NO. 3
MARCH 2017

A PUBLICATION OF THE ENGINEERING SCIENCES SOCIETY
The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

IEICE TRANS. FUNDAMENTALS, VOL.E100–A, NO.3 MARCH 2017

PAPER

Human Wearable Attribute Recognition Using Probability-Map-Based Decomposition of Thermal Infrared Images

Brahmastro KRESNARAMAN[†a)], ***Student Member***, **Yasutomo KAWANISHI**[†], **Daisuke DEGUCHI**[††], **Tomokazu TAKAHASHI**[†††], ***Members***, **Yoshito MEKADA**[††††], **Ichiro IDE**[†], ***Senior Members***, ***and*** **Hiroshi MURASE**[†], ***Fellow***

SUMMARY This paper addresses the attribute recognition problem, a field of research that is dominated by studies in the visible spectrum. Only a few works are available in the thermal spectrum, which is fundamentally different from the visible one. This research performs recognition specifically on wearable attributes, such as glasses and masks. Usually these attributes are relatively small in size when compared with the human body, on top of a large intra-class variation of the human body itself, therefore recognizing them is not an easy task. Our method utilizes a decomposition framework based on Robust Principal Component Analysis (RPCA) to extract the attribute information for recognition. However, because it is difficult to separate the body and the attributes without any prior knowledge, noise is also extracted along with attributes, hampering the recognition capability. We made use of prior knowledge; namely the location where the attribute is likely to be present. The knowledge is referred to as the Probability Map, incorporated as a weight in the decomposition by RPCA. Using the Probability Map, we achieve an attribute-wise decomposition. The results show a significant improvement with this approach compared to the baseline, and the proposed method achieved the highest performance in average with a 0.83 F-score.

key words: *thermal infrared, wearable attribute, recognition, decomposition, probability map*

1. Introduction

Security surveillance systems are very commonly used. Nowadays, the usage is not only limited to commercial buildings, but also residential buildings. These systems play a vital role in the community, as they serve as a deterrent to criminal activity and may be used to help search or identify a criminal. When performing the search or identification, the attributes of that person are important to look at, and these attributes are usually varied. Surveillance cameras generally are located at the ceilings to cover a wider view of the area. However, there are also cases that place cameras in relatively small areas, for example, small alleys, near the entrance to a door, hallways, etc.

In general, surveillance systems employ cameras that work in the visible spectrum; where human eyes see. In this spectrum, illumination is a big factor that needs to be considered. Without sufficient illumination, the captured image might not have enough information which is often vital for the success of any recognition method used. Therefore, surveillance in night-time or in poorly lit areas are challenging tasks. Situations like these make a thermal infrared camera a better option.

A thermal infrared (or simply thermal) camera captures images in thermal infrared spectrum, a portion of the infrared spectrum. A thermal infrared camera produces an image by capturing infrared radiation of an object, therefore, illumination is not an issue. The intensity of a pixel in a thermal infrared image depends on the corresponding temperature of the object. This is the main reason why a thermal infrared camera can be a better alternative in night-time or in poorly lit areas. Figure 1 shows image examples of a person in visible and thermal infrared spectra with and without the presence of wearable attributes.

The term "attribute" itself has a broad meaning. It is defined as a trait or an element of an object. Attributes of a human can be his/her age, gender, race, and so on. For simplicity, these attributes will henceforth be referenced as non-wearable attributes. Meanwhile, this paper focuses on wearable attributes, or objects that people wear. Examples of wearable attributes are glasses, hat, backpack, and so on. Figure 1(b) shows examples where a person is presented with multiple wearable attributes.

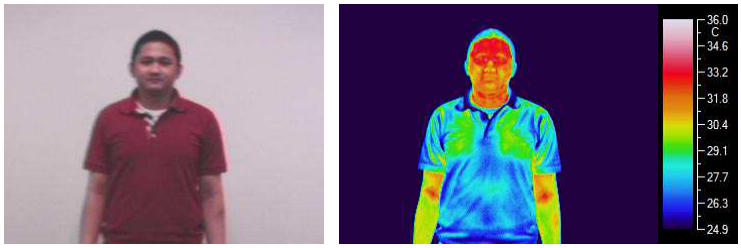
A person shown in Fig. 1(b) is a security risk in places such as banks, because the face is occluded by wearable attributes. Therefore, the information of what wearable attribute a person is wearing is equally important to avoid a compromise in security. This research aims to recognize the wearable attributes of a person in thermal infrared image. This recognition task can be used for searching people based on attributes they are wearing or to warn security about the usage of a certain attribute.

In regards to the recognition of non-wearable attributes, many researches have been done up to now. To name a few, classifications of human expressions [1], [2], race/ethnicity [3], and gender [4], [5]. On the other hand, the researches done to recognize wearable attributes usually include recognition of non-wearable attributes. Kumar et al. [6], [7] utilize various attributes for face verification between two images. Other researches [8]–[11] use images taken from surveillance cameras to find people based on certain attributes. Attributes are also used for person re-

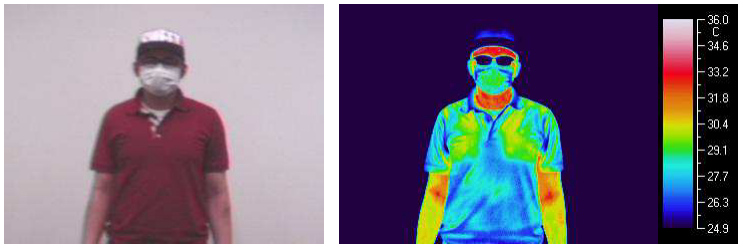
Manuscript received August 1, 2016.
Manuscript revised November 1, 2016.

[†]The authors are with the Grad. School of Information Science, Nagoya Univ., Nagoya-shi, 464-8601 Japan.
[††]The author is with the Information Strategy Office, Nagoya Univ., Nagoya-shi, 464-8601 Japan.
[†††]The author is with the Fac. of Economics and Information, Gifu Shotoku Gakuen Univ., Gifu-shi, 500-8288 Japan.
[††††]The author is with the School of Engineering, Chukyo Univ., Toyota-shi, 470-0393 Japan.
a) E-mail: brahmastrok@murase.m.is.nagoya-u.ac.jp

DOI: 10.1587/transfun.E100.A.854



(a) Without wearable attributes



(b) With wearable attributes (hat, glasses and mask)

Fig. 1 Image examples with and without wearable attributes in visible (left) and thermal infrared spectra (right).

(a) No attribute (b) Mask attribute (c) Backpack attribute

Fig. 2 Example of data with attribute as minority.

identification, for example, works by Layne et al. [12] and Khamis et al. [13].

So far, however, the aforementioned researches are all conducted in the visible spectrum. There are only a small number of attribute recognition studies available in the thermal infrared spectrum. For example, recognition of facial expressions [14], [15] and a small section of eyewear (glasses and sunglasses) detection [9]. To the extent of our knowledge, there are no other works that focus only on wearable attributes in the thermal infrared spectrum.

There are two significant problems to perform recognition of wearable attributes in the thermal infrared spectrum. First, in comparison to the size of human body, the size of attributes are relatively small. Furthermore, intra-class variation of human body is large. Due to these limitations, it is difficult to recognize these attributes in the thermal infrared image. Some works on attribute recognition mentioned previously focus on a smaller region from the image, e.g. the face region.

In this research, we perform recognition by using an image decomposition approach. The decomposition is the key process that produces two images with a size identical to the original. The decomposed images are: one with the wearable attribute information only, and one with the human body without the wearable attributes. We employed the idea of majority and minority to achieve this decomposition, where a certain condition needs to be fulfilled. The condition is that the similarity between the observations needs to be relatively high to serve as the majority. This leads to the situation where the attributes can be separated because they are the minority — a noise — in the data.

To give a clearer idea, assume a collection of thermal infrared images of people. Presented in the collection are human images with no wearable attribute as the majority. If there is a human wearing an attribute in some images, then these attributes will be evaluated as the minority. Figure 2 visualizes this with an example. The mask attribute in Fig. 2(b) and backpack attribute in Fig. 2(c) are considered as the minority in face and torso regions, respectively. Under this assumption, the extraction of the attributes is performed by means of decomposition. Therefore, the method proposes control over the data matrix that will be decomposed.

We have previously proposed the decomposition with Robust Principal Component Analysis (Robust PCA or RPCA) [16] as the basis. RPCA is a modification to Principal Component Analysis (PCA), where the problem of decomposition can be classified according to the approaches developed for it [17]. They are:

- RPCA via Principal Component Pursuit (PCP) [18]
- RPCA via Outlier Pursuit [19]
- RPCA via Iteratively Reweighted Least Squares [20]–[22]
- Bayesian RPCA [23] and its variation [24]
- Approximated RPCA (GoDec [25])

In [16], we constructed the data matrix in such a way that it contains multiple images with no attribute in it as the base data, and one input image that will be either trained or recognized. This arrangement ensures that any attribute is considered as noise by the RPCA and extracted from the input image. Note that the input image may contain multiple, single, or no attributes.

However, there is one drawback that can be perceived in [16]. Since the decomposition was performed without any prior knowledge, there are still some details of clothing and other noise extracted along with the attribute. On the same note, when there are multiple attributes in one image, all of the attributes information is extracted. In this case, the recognition of the desired attribute is hampered by the existence of others.

These problems lead to the idea of an attribute-wise decomposition. By extracting the information only in an attribute-specific region, a more focused decomposition can be achieved. We propose the usage of prior knowledge to tackle the problems. The prior knowledge we utilized is the location where the attribute is likely to be present, which is used as a weight in the decomposition process. Therefore, an attribute-wise decomposition can be achieved. In this paper, we modified RPCA with the introduction of a probability prior.

The usage of a probability prior is very beneficial in situations where there are minimum to no change in viewpoint or angle, for example, near the entrance of a door. In these situations, we can assume that the images contain the upper body region and taken frontally. These images can be obtained from a prior detection step performed on the surveillance video. This step is not covered in our paper, but it should not be difficult, because the presence of human is apparent in the thermal infrared spectrum.

The remaining parts of the paper proceeds in the following. Section 2 elaborates the proposed idea for this research, which is the usage of prior knowledge to achieve an attribute-wise decomposition. Section 3 will then describe the detailed implementation of the proposed method. Experiments and their analyses are explained in Sect. 4, and Sect. 5 concludes the paper.

2. Decomposition by Robust PCA with Probability Map

This section will cover the approach proposed in this paper, which makes use of prior knowledge to achieve an attribute-wise decomposition. However, we first explain in detail the Robust Principal Component Analysis, the basis of our decomposition.

2.1 Robust Principal Component Analysis (RPCA)

Robust Principal Component Analysis (RPCA) is the method used to perform the decomposition. Based on the popular Principal Component Analysis (PCA), RPCA is made to be robust towards corrupted or noisy observations in the data. This is due to the fact that PCA encounters problems when dealing with data that contain outliers and/or noisy observations.

Robust PCA handles the corrupted/noisy observations by separating them from the data. We make use of this decomposition capability for our approach, by considering the wearable attributes as noise instead of a part of the human characteristics. Consequently, wearable attributes in a human image can be decomposed and only the extracted attribute information is used for the recognition.

An idealized version of RPCA was introduced by Candes et al. [18], aimed to decompose a collection of observations M to a low-rank matrix L and a sparse matrix S, as follows:

$$M = L + S, \tag{1}$$

where sparse matrix S contains the noisy/corrupted part of the observation in the data, and its magnitude can be arbitrarily large. Most researches focus on the low-rank components, however, the proposed method utilizes the sparse matrix instead.

There are various techniques to solve RPCA. For this purpose, the proposed method utilizes Principal Component Pursuit (PCP) [18]. Even though some adaptations have been proposed for PCP such as Stable PCP [26] and Local PCP [27], it is still one of the best techniques among its peers [17], [28].

Assume a data matrix M where the observations are represented in column vector. PCP solves the following optimization problem:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \qquad \text{s. t.} \quad L + S = M, \tag{2}$$

where $\|\cdot\|_*$ denotes the nuclear norm, the sum of the singular values given a matrix. $\|\cdot\|_1$ is the l_1-norm with the condition that the matrix is treated as a vector. λ is the balance parameter. The rule of thumb for λ is shown in Eq. (3) where m and n represent the number of rows (dimensions) and columns (observations) of matrix M, respectively.

$$\lambda = \frac{1}{\sqrt{\max(m, n)}} \tag{3}$$

In most of the cases, the value of λ does not need to be adjusted except when prior knowledge is available. PCP is able to obtain the low-rank and the sparse matrices of a data matrix under minimal assumptions. These assumptions are: the low-rank matrix L is not sparse, the sparse matrix S does not have low rank, and the usage of the rule of thumb for the λ value.

2.2 RPCA via Principal Component Pursuit (PCP) with Probability Map

The main idea of the proposed method is to further extract the information of an attribute with prior knowledge, in the region where it likely presents. As an example, to recognize the "glasses" attribute, the only region that we need to check is the region around the eyes. Even if there are "glasses" in another area of the body, for example, in a breast pocket, it is not necessary to recognize them.

Based on this idea, we propose RPCA via Probability Map - Principal Component Pursuit (PM-PCP). The Probability Map is a representation of how likely an attribute will be present in the region of a thermal infrared human image.

As mentioned previously, the proposed method utilizes the sparse matrix and disregards the low-rank representation. The assumption of the decomposition problem with PM is:

$$M = L + P^c \circ S, \tag{4}$$

where P^c is the Probability Map created for attribute c, and $A \circ B$ denotes an element-wise multiplication between matrices A and B. The Probability Map P^c is used as a weight of the sparse matrix in the decomposition, extracting the attribute c where it most likely would be, represented by $P^c \circ S$. At the same time, it penalizes non-attribute information in the region mistakenly extracted along with the attribute information. Therefore, information other than the attribute c is not extracted and left in L. To achieve this, first, the original condition in Eq. (1) is modified with the PM as follows:

$$M = L + (1 - P^c) \circ S + P^c \circ S. \tag{5}$$

Equation (5) contains two terms of the sparse matrix S; the first one represents sparse entries excluding the attribute c and the second one contains the attribute c only. Since we only need the information of attribute c from the second sparse matrix term, L and $(1-P^c)\circ S$ can be grouped together as L', changing the decomposition problem to:

$$M = L' + P^c \circ S. \tag{6}$$

In this case, the L' contains both the low-rank representation and noise from the sparse matrix, information which can be discarded because it is not necessary for recognizing attribute c. On the other hand, $P^c \circ S$ contains the information of attribute c from a specific region based on the Probability Map P^c which will be used later for recognition.

As we can see, the assumption we made earlier in Eq. (4) holds true. Therefore, the original minimization problem by PCP (Eq. (2)) is modified to incorporate the weighting by the Probability Map, shown in Eq. (7).

$$\min_{L', P^c \circ S} \|L'\|_* + \lambda \|P^c \circ S\|_1 \quad \text{s. t.} \quad L' + P^c \circ S = M \tag{7}$$

It is important to note that although the Probability Map is incorporated and hence the minimization problem is modified, the original algorithm used for solving the PCP can still be used with a slight change. The reason is because the PM is constant for each attribute and it does not affect the convergence. The only difference is that there is an element-wise multiplication on the sparse matrix by the Probability Map. Algorithm 1 shows the PM-PCP algorithm in Appendix.

3. Recognition Framework by RPCA via PM-PCP

The most important modification in this framework is the decomposition on a more focused region. This section describes the recognition framework in a more detailed fashion. The process flow of the framework can be seen in Fig. 3.

First, we explain the creation of the Probability Map (PM) which is utilized to customize the decomposition of a specific attribute. Next, the decomposition process by RPCA via PM-PCP is explained and followed by the training and recognition of the classifier. Note that the classifier utilizes the results of the decomposition.

It is important to note that the thermal infrared images utilized in the process uses the "hotblack" color scheme; It is a monochrome color scheme that considers the temperature captured in the pixel, where the hotter the temperature, the closer the pixel value is to zero.

3.1 Creation of Probability Map (PM)

The Probability Map created in this framework is unique to one attribute. In order to create the Probability Map, positive samples of the specific attribute are necessary. These positive samples indicate that the desired attribute is present in the images, otherwise, the images are negative samples.

The data used for the training phase is represented as $D = (\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_K)$, where K is the number of training observations from all attributes. For attribute $c \in \{c_1, c_2, \dots\}$, the positive samples are taken from the training data and represented as $U^c = (\mathbf{u}_1^c, \mathbf{u}_2^c, \cdots, \mathbf{u}_I^c)$ where I is the number of positive observations of attribute c.

Depending on the attribute, one or more rectangular bounding boxes are annotated in the image. For each positive image of attribute c with a size of $e \times f$ [pixels], the probability value of each pixel is $p_i^c(e, f) = 1$ for pixels within the rectangular regions and $p_i^c(e, f) = 0$ otherwise, where $i = 1, 2, \dots, I$. After the annotation is performed on all of the positive images, the probability of attribute c per pixel are calculated by:

$$p^c(e, f) = \frac{1}{I} \sum_{i=1}^{I} p_i^c(e, f). \tag{8}$$

The Probability Map p^c is a pixel map with a size of

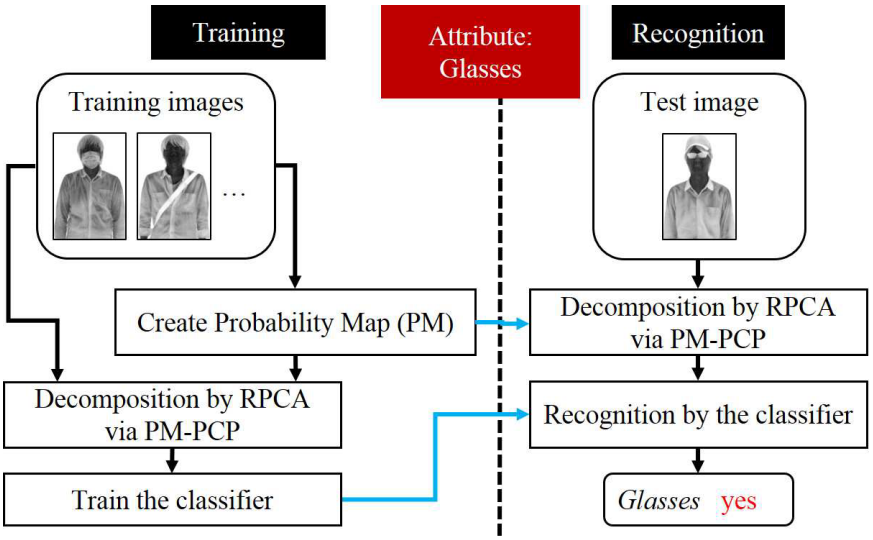
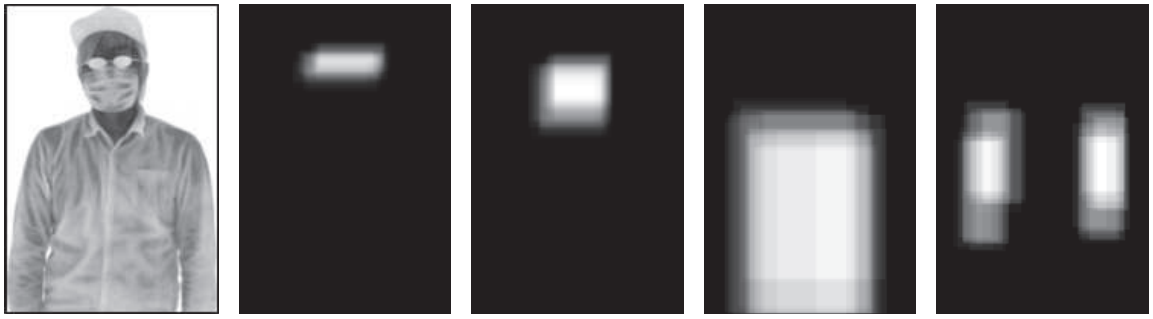


Fig. 3 Process flow of the recognition framework.



(a) Thermal image (b) Glasses PM (c) Mask PM (d) Shoulder bag PM (e) Backpack PM

Fig. 4 Example of thermal image and various Probability Maps (PMs).

$e \times f$ [pixels] of which the pixels value is the probability calculated in Eq. (8) for attribute c, ranging from 0 to 1. In order to use the PM, p^c needs to be rearranged to a column vector $\mathbf{p}^c$ and duplicated n times so it has the same dimension as M, which is $m \times n$. As mentioned previously, m represents the number of rows (dimensions), which is $m = e \times f$, and n is the number of observations. This is represented by $P^c = (\mathbf{p}_1^c, \mathbf{p}_2^c, \cdots, \mathbf{p}_n^c)$. Figure 4 shows examples of Probability Maps of various attributes.

3.2 Decomposition and Recognition Process

In the beginning of the decomposition process, images that have no wearable attribute are categorized in vector form as base data and represented as $B = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_J)$ with J observations. The base data is used both in the training and the recognition phases. Each phase is explained further in their respective subsections.

3.2.1 Training Phase

The training data of all attributes $D = (\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_K)$ introduced previously are divided into multiple training data per attribute $Q^c = (\mathbf{q}_1^c, \mathbf{q}_2^c, \cdots, \mathbf{q}_H^c)$ where $H \leq K$. Note that Q^c has both positive and negative samples of attribute c. This attribute-specific training data Q^c will then be decomposed. Therefore, for the training of attribute c, PM-PCP is used to solve:

$$\begin{bmatrix} B & \mathbf{q}_h^c \end{bmatrix} = M_h = L'_h + (P^c \circ S)_h, \tag{9}$$

where

$$(P^c \circ S)_h = \begin{bmatrix} \mathbf{s}_1^c & \cdots & \mathbf{s}_J^c & \mathbf{x}_h^c \end{bmatrix}. \tag{10}$$

The entry $\mathbf{x}_h^c$ is the sparse representation that corresponds to $\mathbf{q}_h^c$. The decomposition process of training data is performed H times, and the sparse results are grouped together as $X^c = (\mathbf{x}_1^c, \mathbf{x}_2^c, \cdots, \mathbf{x}_H^c)$ for the training of the classifier.

Since the training data $X^c = (\mathbf{x}_1^c, \mathbf{x}_2^c, \cdots, \mathbf{x}_H^c)$ contains positives and negatives of attribute c, any method can be used as the classifier in this recognition framework. Note that one classifier is trained for each attribute, and in this case, we trained for attribute c. As an example, for a binary classifier, we provide both the positive and negative samples of attribute c until it is capable to classify them. When the classifier is trained, it marks the end of the training phase.

3.2.2 Recognition Phase

To be able to recognize attribute c in an input thermal image $\mathbf{t}$, PM-PCP needs to solve the following:

$$\begin{bmatrix} B & \mathbf{t} \end{bmatrix} = M_t = L'_t + (P^c \circ S)_t, \tag{11}$$

where

$$(P^c \circ S)_t = \begin{bmatrix} \mathbf{s}_1^c & \cdots & \mathbf{s}_J^c & \mathbf{y}^c \end{bmatrix}. \tag{12}$$

Note that Eq. (12) includes an entry $\mathbf{y}^c$, the sparse representation of $\mathbf{t}$. This entry is used as the input for the classifier, which has been trained for attribute c. The result of the recognition test shows whether the attribute c is present in $\mathbf{t}$ or not. For example, in the case of a binary classifier, the output will be 1 when the attribute c is present, and 0, otherwise.

4. Experiments

In this section, we will present the experiments conducted to evaluate the capability of the proposed method. First, we explain the dataset that is used, including various attributes present. Then, we describe details of the experimental setup, as well as the experiments. The results and analysis are provided at the end.

4.1 Dataset

We used a dataset from [16], a private dataset we created for this research, since a publicly available dataset that contains various attributes in thermal images and details of the data capture settings is hard to come by. The dataset contains a total of 408 frontal thermal infrared images from fourteen people, with up to seven different wearable attributes per person. The attributes available in the dataset and the number of positives available can be seen in Table 1.

We used AVIO's TVS-500EX camera [29] for the creation of the dataset; a camera that has the capability to capture both thermal and visible images simultaneously. The size of the image captured in thermal infrared is 320×240 pixels.

Table 1 Distribution of the seven wearable attributes in the dataset.

Attributes	# of images
No attribute	28
Glasses	168
Mask	168
Hat	112
Helmet	80
Hoodie	40
Shoulder bag	48
Backpack	40

The camera can capture infrared wavelength that ranges from 8 to 14 μm.

The images were taken indoors at room temperature (around 22–25° Celsius), while the camera captures infrared radiation whose temperature ranges from 25 to 36° Celsius. Subjects were standing around 3.0 to 3.5 m from the camera, and only the upper region of the body was captured by the thermal infrared camera. As shown in Fig. 5, images in the dataset are using the "hotblack" color scheme. This means the closer the pixel is to 0, the hotter the temperature is. We cropped the human body regions manually as part of the preprocessing. The average size of the cropped region is 140×204 pixels.

4.2 Experimental Setup

To evaluate the capability of the proposed method, we compared the recognition performance. The classifier of choice for this experiment was Support Vector Machine (SVM) [30], with the decomposition performed on a pixel value. The experiment was conducted with a leave-one-person-out cross-validation. The way the data was divided ensures the training and test data do not intersect.

As a comparison, we perform recognition using other methods, as follows:

- No Decomposition: Thermal images are directly used for training and testing of the classifier without the decomposition. This method serves as the baseline.
- Conventional Average: First, we calculate the average from the base data $B = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_J)$, and represent it as $\bar{\mathbf{b}}$. This average serves the same way as the base data, to "decompose" both the training and the test data. The "decomposition" is performed by subtracting $\bar{\mathbf{b}}$ from the input, instead of using the PM-PCP.
- Conventional PCA: Principal Component Analysis (PCA) is utilized to obtain the low rank version of test data $\mathbf{t}$. From the base data $B = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_J)$, the projection matrix V is calculated by PCA. The test data $\mathbf{t}$ is projected to and back from the Eigenspace, represented by $\tilde{\mathbf{t}}$. The "decomposition" is performed between the low rank version $\tilde{\mathbf{t}}$ and the test data $\mathbf{t}$ itself by means of subtraction.
- RPCA via PCP: The decomposition method proposed in our previous work [16].
- Post-PCP PM: First, the decomposition is performed just like in [16]. After the decomposition, the Probability Map is applied on the sparse entry that corresponds to the input. This is a simpler version of using the Probability Map without modifying the RPCA via PCP problem.

For each attribute, both positive and negative samples are fed to all of the methods. The decomposition results are forwarded to the SVM. Let us assume we want to recognize "glasses" attribute. Therefore, the positive samples are images where "glasses" is present and the negative samples are images where "glasses" is not present. When SVM success-

(a) No attribute (b) Mask attribute (c) Glasses, mask and hat attributes (d) Backpack attribute

Fig. 5 Image examples from the dataset.

Table 2 Recognition results of the various methods evaluated in F-Score. Bold values mark the highest value in their respective column.

Methods	Attributes						Average over attributes in:		
	Glasses	Mask	Hat	Helmet	Shoulder bag	Backpack	Head region	Torso region	All regions
No Decomposition	0.607	0.603	0.471	0.357	0.222	0.200	0.51	0.21	0.41
Conventional Average	0.834	0.988	0.551	0.800	0.538	0.849	0.79	0.69	0.76
Conventional PCA	0.854	0.985	0.476	0.645	0.698	0.892	0.74	0.79	0.76
RPCA via PCP [16]	0.861	0.982	0.502	0.754	**0.723**	**0.947**	0.77	**0.84**	0.79
Post-PCP PM	0.918	**0.997**	0.581	0.843	0.675	0.841	0.83	0.76	0.81
RPCA via PM-PCP (proposed)	**0.919**	0.991	**0.623**	**0.868**	0.692	0.868	**0.85**	0.78	**0.83**

fully recognizes "glasses" from a positive sample (output 1), it is a True Positive case. A True Negative case happens when SVM recognizes that there is no "glasses" from a negative sample (output 0). Based on this binary output, precision and recall can be calculated, and consequently, F-score. The F-score is used to evaluate the performance of each method.

4.3 Results and Analysis

Table 2 shows the recognition results of the experiment. We provide results of each attribute and average over attributes on the head region, torso region, and as a whole. Compared to the baseline, the approach of performing decomposition on the thermal infrared images shows a significant performance increase.

Unexpectedly, the results of both Conventional Average and Conventional PCA methods are relatively close to the decomposition by RPCA via PCP. Overall, the decomposition via PM-PCP achieves the highest performance with an average of 0.83 F-score, outperforming all the other methods. Coming in second is the Post-PCP PM method, which confirms the usefulness of Probability Map for the decomposition. Other than the mask attribute, it fails to outperform the RPCA via PM-PCP method.

However, the RPCA via PM-PCP method is not the best for attributes in the torso region, as the decomposition by RPCA via PCP and Conventional PCA performs better. Except for the backpack attribute where the RPCA via PCP holds a clear advantage, the differences between the three methods are relatively small.

The most likely reason is due to the variations of human body (mainly the height) in the dataset which in turn makes the rectangular region for the attribute labeling not as accurate. The by-product of this is that the probability of some parts in the backpack PM are relatively low, and the weighted decomposition makes the sparse representation of the attribute information of those parts not as apparent. This can be observed in the bottom row of Fig. 6, where the qualitative comparison is shown. Note that the resulting images have been normalized for visualization purpose. The normalization is performed per image, where minimum and maximum pixel values are changed to 0 and 1, respectively.

In Fig. 6, it can also be seen that the sparse representation of the thermal infrared image clearly shows the extracted attributes decomposed by each method. The Conventional Average produces a lot more noise in the results than other decomposition methods, where the body shape can still be clearly seen. The Conventional PCA shows better decomposition results than the Conventional Average, although not by much. The results of the decomposition via PCP are better than the previous methods, although some clothes and hair details are still noticeable.

The results of the Post-PCP PM method show a more focused decomposition. However, there is still noise unrelated to the attribute we try to recognize. The proposed method; decomposition via PM-PCP, successfully extracts the attribute information almost exclusively. The usage of Probability Map proves to be helpful for the decomposition, as shown in both quantitative and qualitative results. Although the results of the decomposition by RPCA via PM-

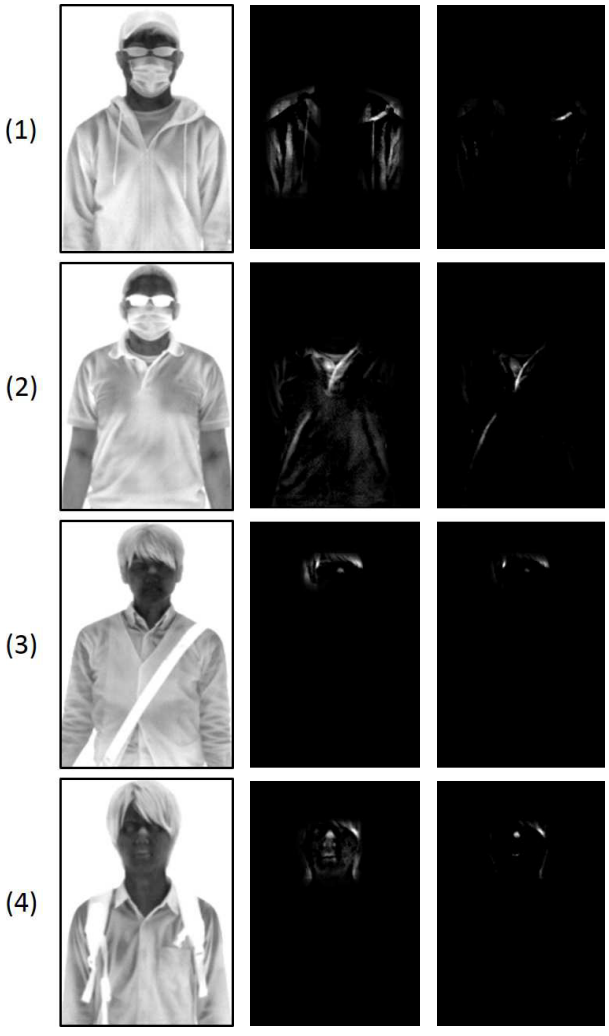
Fig. 6 Examples of decomposition results by various methods. The decomposition is made for: (1) "glasses", (2) "mask", (3) "shoulder bag", and (4) "backpack" attributes. For visualization purpose, the minimum and maximum pixel values for each image have been normalized to 0 and 1, respectively.

PCP has some erosion on the attributes, the Post-PCP PM method contains more non-attribute noise, which is detrimental to the recognition performance.

For discussion, we provide results of decomposition by methods that use PM in the case of negative sample used as the input image. These are the cases where the decomposed image does not contain the attribute we are trying to recognize. For example, let's consider a case where the input image contains only the "backpack" attribute when we performed the decomposition for the "glasses" attribute. Figure 7 shows some of the results of such decompositions. Note that the decomposition results are normalized just like before, i.e. minimum and maximum of each image are normalized to 0 and 1, respectively. It can be seen that there are details of clothing and face in the decomposition results of both methods. However, the proposed method produces better results with less noise than Post-PCP PM method. Post-PCP PM focuses on the area corresponding to the PM, but more clothes and face details are taken. As mentioned previously, this leads to lower recognition results than the proposed method.

5. Conclusion

This paper addressed the attribute recognition problem in thermal infrared images, specifically on the wearable attribute. Attribute recognition as a whole is a growing field of research, but only a few studies have been conducted in



(a) Original image (b) Post-PCP PM (c) RPCA via PM-PCP

Fig. 7 Examples of decomposition results by methods that use PM in the case of negative input images. The decomposition is made for: (1) "backpack", (2) "shoulder bag", (3) "glasses", and (4) "mask" attributes. For visualization purpose, the minimum and maximum pixel values for each image have been normalized to 0 and 1, respectively. Since the images do not have the target attributes, pixels of the decomposed images are close to black.

thermal infrared spectrum.

The purpose of this research was to identify a specific wearable attribute of a person in a thermal infrared image. We adopted an attribute-wise decomposition approach to this problem, extracting the attributes information with prior knowledge and using it for recognition. The proposed method employs a decomposition by Robust PCA via PM-PCP, a modification of PCP used in our previous work [16]. We introduced Probability Map (PM) to achieve a more focused decomposition.

The results of experiments performed showed a significant increase in performance with the decomposition approach. Decomposition by Robust PCA via PM-PCP achieved the highest performance in average with a 0.83 F-score, outperforming all other methods. By visual observation, the resulting images also showed relatively better extracted attributes.

Increasing the size of the dataset is important for further research, which could provide both more intra- and inter-class variations. Immediate attention can also be given to the alignment of the body in the dataset. In the case of the creation of the Probability Map, instead of making a rectangular region, a more detailed — pixel-wise — labeling of an attribute can potentially be helpful. Lastly, it is desirable to perform recognition on images taken in a real-world situation.

Acknowledgement

Parts of this research were supported by MEXT, Grant-in-Aid for Scientific Research. Authors would like to thank the members and former members of Murase Laboratory, Nagoya University for their participation in the creation of the dataset.

References

[1] M.S. Bartlett, G. Littlewort, I. Fasel, and J.R. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction," Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition 2003 Workshops, pp.53–59, Madison, Wisconsin, USA, June 2003. DOI:10.1109/CVPRW.2003.10057

[2] M.J. Lyons, J. Budynek, A. Plante, and S. Akamatsu, "Classifying facial attributes using a 2-D Gabor wavelet representation and discriminant analysis," Proc. 4th IEEE Int. Conf. Automatic Face and Gesture Recognition, pp.202–207, Grenoble, France, March 2000. DOI:10.1109/AFGR.2000.840635

[3] G. Shakhnarovich, P.A. Viola, and B. Moghaddam, "A unified learning framework for real time face detection and classification," Proc. 5th IEEE Int. Conf. Automatic Face and Gesture Recognition, pp.14–21, Washington, DC, USA, May 2002. DOI:10.1109/AFGR.2002.1004124

[4] B.A. Golomb, D.T. Lawrence, and T.J. Sejnowski, "Sexnet: A neural network identifies sex from human faces," Advances in Neural Information Processing Systems 3, pp.572–577, Denver, CO, USA, Nov. 1990.

[5] S. Wang, Z. Gao, S. He, M. He, and Q. Ji, "Gender recognition from visible and thermal infrared facial images," Multimed. Tools Appl., vol.75, no.14, pp.8419–8442, July 2016. DOI:10.1007/s11042-015-2756-5

[6] N. Kumar, A.C. Berg, P.N. Belhumeur, and S.K. Nayar, "Attribute and simile classifiers for face verification," Proc. 12th IEEE Int. Conf. Computer Vision, pp.365–372, Kyoto, Japan, Sept. 2009. DOI:10.1109/ICCV.2009.5459250

[7] N. Kumar, A.C. Berg, P.N. Belhumeur, and S.K. Nayar, "Describable visual attributes for face verification and image search," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.33, no.10, pp.1962–1977, Oct. 2011. DOI:10.1109/TPAMI.2011.48

[8] R. Feris, R. Bobbitt, L. Brown, and S. Pankanti, "Attribute based people search: Lessons learnt from a practical surveillance system," Proc. 4th ACM Int. Conf. Multimedia Retrieval, pp.153–160, Glasgow, Scotland, UK, April 2014. DOI:10.1145/2578726.2578732

[9] D.A. Vaquero, R.S. Varis, D. Tran, L. Brown, A. Hampapur, and M. Turk, "Attribute based people search in surveillance environment," Proc. IEEE Computer Society Workshop on Applications of Computer Vision 2009, pp.1–8, Snowbird, UT, USA, Dec. 2009. DOI:10.1109/WACV.2009.5403131

[10] J. Thornton, J. Baran-Gale, D. Butler, M. Chan, and H. Zwahlen, "Person attribute search for large-area video surveillance," Proc. IEEE Int. Conf. Technologies for Homeland Security 2011, pp.55–61, Waltham, MA, USA, Nov. 2011. DOI:10.1109/THS.2011.6107847
[11] Y. Deng, P. Luo, C.C. Loy, and X. Tang, "Pedestrian attribute recognition at far distance," Proc. 22nd ACM Int. Conf. Multimedia, pp.789–792, New York, NY, USA, Nov. 2014. DOI:10.1145/2647868.2654966
[12] R. Layne, T.M. Hospedales, and S. Gong, "Person re-identification by attributes," Proc. 23rd British Machine Vision Conf., pp.24.1–24.11, Guildford, England, UK, Sept. 2012. DOI:10.5244/C.26.24
[13] S. Khamis, C.H. Kuo, V.K. Singh, V.D. Shet, and L.S. Davis, "Joint learning for attribute-consistent person re-identification," Proc. 13th European Conf. Computer Vision, pp.134–146, Zurich, Switzerland, Sept. 2014. DOI:10.1007/978-3-319-16199-0_10
[14] B. Hernandez, G. Olague, R. Hammoud, L. Trujillo, and E. Romero, "Visual learning of texture descriptors for facial expression recognition in thermal imagery," Computer Vision and Image Understanding, vol.106, no.2–3, pp.258–269, May–June 2007. DOI:10.1016/j.cviu.2006.08.012
[15] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez, "Automatic features localization in thermal images for facial expression recognition," Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition 2005 Workshops, pp.14–21, San Diego, CA, USA, June 2005. DOI:10.1109/CVPR.2005.415
[16] B. Kresnaraman, Y. Kawanishi, T. Takahashi, D. Deguchi, Y. Mekada, I. Ide, and H. Murase, "Human wearable attribute recognition using decomposition of thermal infrared images," Proc. 22nd Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp.123–127, Takayama, Japan, Feb. 2016.
[17] T. Bouwmans and E.H. Zahzah, "Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance," Computer Vision and Image Understanding, vol.122, pp.22–34, May 2014. DOI:10.1016/j.cviu.2013.11.009
[18] E.J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," J. ACM, vol.58, no.3, p.11, May 2011. DOI:10.1145/1970392.1970395
[19] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," Advances in Neural Information Processing Systems 23, pp.2496–2504, Vancouver, BC, Canada, Dec. 2010. DOI:10.1109/TIT.2011.2173156
[20] C. Guyon, T. Bouwmans, and E.H. Zahzah, "Foreground detection via robust low rank matrix decomposition including spatio-temporal constraint," Proc. 11th Asian Conf. Computer Vision Workshops, pp.315–320, Daejeon, Korea, Nov. 2012. DOI:10.1007/978-3-642-37410-4_28
[21] C. Guyon, T. Bouwmans, and E. Zahzah, "Foreground detection via robust low rank matrix factorization including spatial constraint with iterative reweighted regression," Proc. 21st APR Int. Conf. Pattern Recognition, pp.2805–2808, Tsukuba, Japan, Nov. 2012.
[22] C. Guyon, T. Bouwmans, and E.H. Zahzah, "Moving object detection via robust low rank matrix decomposition with IRLS scheme," Proc. 8th Int. Symposium Visual Computing, pp.665–674, Rethymnon, Crete, Greece, July 2012. DOI:10.1007/978-3-642-33179-4_63
[23] X. Ding, L. He, and L. Carin, "Bayesian robust principal component analysis," IEEE Trans. Image Process., vol.20, no.12, pp.3419–3430, Dec. 2011. DOI:10.1109/TIP.2011.2156801
[24] S.D. Babacan, M. Luessi, R. Molina, and A.K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," IEEE Trans. Signal Process., vol.60, no.8, pp.3964–3977, Aug. 2012. DOI:10.1109/TSP.2012.2197748
[25] T. Zhou and D. Tao, "Godec: Randomized low-rank & sparse matrix decomposition in noisy case," Proc. 28th Int. Conf. Machine Learning, pp.33–40, Washington, DC, USA, June 2011.
[26] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," Advances in Neural Information Processing Systems 22, pp.2080–2088, Vancouver, BC, Canada, Dec. 2009.
[27] B. Wohlberg, R. Chartrand, and J. Theiler, "Local principal component pursuit for nonlinear datasets," Proc. 37th IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp.3925–3928, Kyoto, Japan, March 2012. DOI:10.1109/ICASSP.2012.6288776
[28] C. Guyon, E.H. Zahzah, and T. Bouwmans, "Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis," InTech, March 2012. DOI:10.5772/38267
[29] "Thermal video system advanced thermo TVS-500EX," http://www.infrared.avio.co.jp/en/products/ir-thermo/lineup/tvs-500ex/spec.html. Accessed Oct. 20, 2016.
[30] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol.20, no.3, pp.273–297, Sept. 1995. DOI:10.1007/BF00994018
[31] S. Kontogiorgis and R.R. Meyer, "A variable-penalty alternating directions method for convex optimization," Math. Program., vol.83, no.1, pp.29–53, Jan. 1998. DOI:10.1007/BF02680549

Appendix: RPCA via PM-PCP Algorithm

The algorithm used to solve the minimization problem of RPCA via PM-PCP is based on the original PCP with a slight change, which is an inclusion of an element-wise matrix multiplication. Since the Probability Map is constant, the usage of PM does not affect convergence. The original PCP algorithm is solved by alternating directions method [31], a special case of Augmented Lagrange Multiplier (ALM) algorithms. More details of the original PCP algorithm can be seen in [18].

Algorithm 1 RPCA via PM-PCP Algorithm

1: **initialize:** $S_0 = Y_0 = 0, \mu > 0$
2: **while** not converged **do**
3: $\quad L_{k+1} = D_{\mu^{-1}}(M - S_k + \mu^{-1}Y_k)$
4: $\quad S_{k+1} = P^c \circ S_{\lambda\mu^{-1}}(M - L_{k+1} + \mu^{-1}Y_k)$
5: $\quad Y_{k+1} = Y_k + \mu(M - L_{k+1} - S_{k+1})$
6: **end while**
7: **output:** L, S

Algorithm 1 shows the PM-PCP algorithm, where $\mu = n_1 n_2 / 4||M||_1$ (note that n_1, n_2 are the size of M) and Y is the Lagrange multiplier matrix. $D(X) = US(\Sigma)V^*$ denotes the singular value thresholding operator, where $X = U\Sigma V^*$ is singular value decomposition. $A \circ B$ is an element-wise matrix multiplication. The convergence is reached when $||M - L - S||_F \leq \delta ||M||_F$, where $\delta = 10^{-7}$ in the experiment.

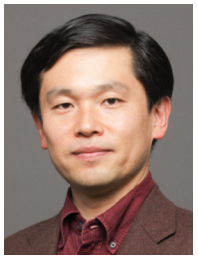
Brahmastro Kresnaraman received his BCS from University of Indonesia and MSCS from Chukyo University, Japan in 2010 and 2014, respectively. He is currently pursuing his Doctoral degree in Nagoya University, Japan. His research interests include human image processing and attribute classification, especially in thermal infrared spectrum.

Yasutomo Kawanishi received his B.Eng. and M.Eng. degrees in Engineering and a Ph.D. degree in Informatics from Kyoto University, Japan, in 2006, 2008, and 2012, respectively. He became a Post Doctoral Fellow at Kyoto University in 2012. He moved to Nagoya University, Japan as a Designated Assistant Professor in 2014. Since 2015, he has been an Assistant Professor at Nagoya University, Japan. His research interests are pedestrian-centric vision, which includes pedestrian detection, tracking, and retrieval, for surveillance and in-vehicle videos. He received the best paper award from SPC2009, and Young Researcher Award from IEEE ITS Society Nagoya Chapter. He is a member of IEEE.

Daisuke Deguchi received his B.Eng. and M.Eng. degrees in Engineering and a Ph.D. degree in Information Science from Nagoya University, Japan, in 2001, 2003, and 2006, respectively. He is currently an Associate Professor at Information Strategy Office, Nagoya University, Japan. He is working on the object detection, segmentation, recognition from videos, and their applications to ITS technologies, such as detection and recognition of traffic signs. He is a member of IEEE.

Tomokazu Takahashi received his B.Eng., M.Eng., and Ph.D. degrees from Ibaraki University, Japan, in 1997, 2000, and 2003, respectively. He became a Post Doctoral Researcher for the 21st Century COE program at Nagoya University from 2003, and a JSPS Post Doctoral Research Fellow from 2005. Since 2008, He has been an Associate Professor at Gifu Shotoku Gakuen University, Japan. From 2013 to 2014 he was a visiting research scientist at Columbia University, New York. He is working on fundamental research of image recognition and its applications.

Yoshito Mekada received the B.S., M.S., and Ph.D. degrees in information engineering from Nagoya University, Japan in 1991, 1993 and 1996, respectively. He is currently a Professor at the School of Engineering, Chukyo University, Japan. His current research interests include medical three dimensional image processing, computer vision, and their applications. He is a member of Japanese Society of Medical Imaging Technology and IEEE.



Ichiro Ide received his B.Eng., M.Eng., and Ph.D. degrees from The University of Tokyo, Japan in 1994, 1996, and 2000, respectively. He became an Assistant Professor at the National Institute of Informatics, Japan in 2000. Since 2004, he has been an Associate Professor at Nagoya University, Japan. He had also been a Visiting Associate Professor at National Institute of Informatics, Japan from 2004 to 2010, an Invited Professor at Institut de Recherche en Informartique et Systèmes Aléatoires (IRISA), France in 2005, 2006, and 2007, a Senior Visiting Researcher at ISLA, Instituut voor Informatica, Universiteit van Amsterdam, the Netherlands from 2010 to 2011. His research interest ranges from the analysis and indexing to retargeting of multimedia contents, especially in large-scale broadcast video archives, mostly on news, cooking, and sports contents. He is a senior member of IPS Japan, and members of ITE, JSAI, IEEE, and ACM.

Hiroshi Murase received the B.Eng., M.Eng., and Ph.D. degrees in Electrical Engineering from Nagoya University, Japan. In 1980 he joined the Nippon Telegraph and Telephone Corporation (NTT). From 1992 to 1993 he was a visiting research scientist at Columbia University, New York. From 2003 he is a Professor at Nagoya University, Japan. He was awarded the IEICE Shinohara Award in 1986, the Telecom System Award in 1992, the IEEE CVPR (Conference on Computer Vision and Pattern Recognition) Best Paper Award in 1994, the IPS Japan Yamashita Award in 1995, the IEEE ICRA (International Conference on Robotics and Automation) Best Video Award in 1996, the Takayanagi Memorial Award in 2001, the IEICE Achievement Award in 2002, the Ministry Award from the Ministry of Education, Culture, Sports, Science and Technology in 2003, and the Medal with Purple Ribbon in 2012. Dr. Murase is IEEE Fellow, and a member of the IPS Japan.