Garment Detection by iterating the Grabcut using the most reliable Seed

Nozomi matsuda ^{a,*}, Lifeng Zhang ^a

^a Kyushu Institute of Technology, 1-1 Sensui-cho Tobata-ku Kitakyusyu-shi Fukuoka-ken, 8048550, Japan

* p349432n@mail.kyutech.jp

Abstract

The technology using a garment image is applied to various researches and systems. The accuracy of the technology based on a garment image depends on a result of a garment detection method. However, almost existing methods are executed in an environment of the person standing or appearing clearly as assumption. Therefore they may not detect a garment in a challenging environment. A flexible method is required for the reason. In this research, a garment detection method is proposed to detect it in that situation. The method iterates the Grabcut using the most reliable seed. By evaluating the areas detected by Grabcut repeatedly, it decides the seed and iterates the detection. From an examination, it is confirmed that it detects the garment in the challenging environment.

Keywords: Garment detection, Grabcut, Seed definition.

1. Introduction

The technology using a garment image is applied to various researches and systems. In research field, it is adopted to person identification, person tracking, occupation predicting and others. Because characteristics of garment (shape, color, pattern and others) have diversity, it is suitable to identification and recognition methods. On the other hand, the accuracy of the technology based on a garment image depends on a result of the garment detection method. Therefore, the detection method is still being researched.

The garment detection methods are proposed based on several perspective. These are mainly divided into three methods. There are graph based methods, shape based method and model based method. We brief each method with a few sub-section.

1.1 Graph based method

The graph based method aims to divide an image to a

foreground and a background. In order to complete the aim, it uses a color space distribution and location of each pixel. It is able to detect a garment by a prediction based on a value calculated from each factor. However, before conducting the method, it needs to roughly allocate labels (usually foreground or background) to each pixel. In order to meet the requirement, the labels are often decided based on other absolute coordinates like a face. Furthermore, it is difficult to divide an image into foregrounds and backgrounds, if two color space distributions are similar.

1.2 Shape based method

The shape based method aims to detect a shape which was learned from objective features in advance. Usually, it needs a lot of image to learn the shape. Therefore, it is able to detect a garment by learning garment shapes as an objective feature. However, the garment shape is not constant as we know. Therefore, this method may be not suitable if it does not limit a garment condition.

1.3 Model based method

The model based method aims to detect a garment by combining a few other techniques ⁽¹⁾. First one is the human pose estimation method ⁽²⁾ ⁽³⁾ and second one is super-pixel method.

The human pose estimation method uses a feature called a model. The model is constructed by learning from a lot of human image. By the learning, it is able to detect a human and estimate one's pose of upper-body, whole-body, and others in an image. In case of garment detection, it is important to estimate a torso by this way. The Fig 1 shows a few cases of pose estimation. A red line indicates a torso position. Other color lines indicate other body parts. The (a) and (d) are pieces of this research images. The left image shows a torso correctly. However, the right image shows an improper torso clearly. This is a result of estimating a pose of a person standing back of a person wearing orange

Fig. 1. The results of pose estimation method ⁽²⁾.

garment. In this case, a garment may be not detected correctly.

The super-pixel method unifies each pixel within an image. The unified pixels are called a super-pixel. Each pixel within a super-pixel has similar characteristic from viewpoints of a color space distribution and its location. By executing the method, various objects (like a face, hair, a body, hands and others) appear as one of super-pixels.

The model based method uses the estimated human pose and the super-pixel. It extracts super-pixels corresponding to an upper-body which is estimated from human pose. The extracted super-pixels are defined as a garment. The model based method is assumed to be suitable to detect a garment because these core methods are sophisticated. However, the human pose estimation method absolutely estimates a pose of an objective. If it must estimate an upper-body pose despite an upper-neck appears within an image, it obtains an improper pose like the Fig 1(b). Therefore, this method also requires to limit a condition of an objective or select appropriate model.

1.4 Person identification system using garment feature

Currently, we are researching to construct a system which monitors kindergarten children and identifies each child ⁽⁴⁾. For the identification method, two garment features are adopted in the research. From previous research, these features were assumed to be effective to identify person. On the other hand, it was also discovered that the detecting a garment from our challenging image is difficult. The images contain over one child and complex background. Furthermore, a posture of the child is not constant and appearing body is also not constant because the body can be hidden by furniture or other child. For these reasons, graph based method is assumed to be better than other kind of method. However, it was confirmed that existing graph based method couldn't detect garment enough. Therefore, an effective garment detection method is required. In this research, we propose one method to attain the objective. This

is one of methods using Grabcut ⁽⁵⁾ technique.

2. Related work

2.1 Grabcut

The Grabcut is representative graph based method. Although this method was proposed as interactive image segmentation method, it is adopted in various automatic methods because of its simple theory and high accuracy. It able to detect an object by being given a part of objective area called seed. It creates two probability models of color space distribution regarding the seed and the non-seed, using Gaussians Mixtures Model (GMM). These probability models include some components (usually 5) of GMM. After creating the models, it assigns each pixel to the foreground and the background by basing on these models and location of pixels.

Concretely, the Grabcut considers object detection as energy minimization problem. It calculates the $E(\cdot)$ according to the

$$E(p, l, d, m) = U(p, l, m, d) + V(p, l) \quad (1)$$

The $p \in \{0, \dots, N (= \text{width} * \text{height of an image})\}$ means each pixel. The $l = 0$ or 1 , is a label that means each pixel is assigned to the background or the foreground. The $m \in \{m_0, \dots, m_N\}$ means an unique GMM component corresponding to each pixel. The d means coefficient of each GMM. The $U(\cdot)$ is called data term. It means likelihood of that each pixel will be assigned to a label given in advance. The $V(\cdot)$ is called smoothing term. Its score will become larger if adjacent pixels have been assigned to different label. The Grabcut calculates any combinations of pixels and labels to minimize the $E(\cdot)$. Therefore, the combination indicating the smallest $E(\cdot)$ is assumed the best condition that object is detected well from an image. Furthermore, it is able to raise the precision of detection by re-assigning labels based on the result of $E(\cdot)$ to each pixel as new seed.

2.2 Simplified garment detection method

As basic method to detect a garment by using the Grabcut, one way that assigns a seed to some pixels locating under a face exists. This one supposes that a garment exists under a face. Fig 2 shows a few result. The left image shows an original image and the right image shows a result image. In Fig 2, a rectangle locating under the face was assigned as a first seed. Its size is same to a face size.

In our previous research, some pixels within an area locating just under a face were assigned to a seed, and the

(a)

(b)

Fig. 2. The results of simplified garment detection ⁽⁶⁾.

result which is similar to Fig 2 appeared. The first seed size is 4 times that of face area. This way detects garment enough if a child stands and nothing interrupts to front of the child. However, that condition is an exceedingly limited scene for children staying at the kindergarten. Therefore, this is not suitable for our objective.

2.3 Garment detection method by voting

As described at section 2.2, it is not able to detect garment well by using the simplified method. A flexible method is required to be adapted to the various posture, so one method by voting ⁽⁶⁾ was proposed.

It executes the Grabcut to some areas within a fixed area. Then it votes to any pixel coordinates within the detected areas. Finally, the pixels indicating high voting score than average voting score are decided as garment area. This method supposes that an expected area as objective will often appears even if the Grabcut is executed using some different seeds. However, the method couldn't output a promising result in our research. Fig 3 shows a few result. The left image shows an original image and the right image shows a result image. In the result, some white areas were detected as a garment. In Fig 3, the truth garments contain marks and the background is complicated, so the detection didn't work well.

3. Proposed method

In this research, a garment detection method by iterating

(a)

(b)

Fig. 3. The results of garment detection method ⁽⁶⁾.

the Grabcut using the most reliable seed is proposed. The proposed method supposes that a garment exists under a neck. Furthermore, it is assumed that garment detection from challenging images is possible by executing the Grabcut at three locations around the neck. This is divided into four procedures. The Fig 4 shows the flowchart of the method. Firstly, it detects faces in an image and clips areas including each person. Secondly, it detects a neck. Thirdly, it executes area detection to create a seed for the garment detection. Finally, it detects a garment by using an area which detected as a seed in former procedure. Before the explanation of each procedure, we describe about the iterating the Grabcut using the most reliable seed because it is executed in almost procedures.

3.1 Iterating the Grabcut using the most reliable seed

Generally, although several objective areas are detected after executing the Grabcut, only one area is required because real objective areas are unique in our research. Therefore the most reliable area must be decided according to some theories. Furthermore, by assigning the area as new seed repeatedly, it is assumed to detect an objective well. The Fig 5 shows the flowchart of the method.

(1) Inputting an image

This process receives an image including an objective from other method.

(2) Assigning pixels as first seed

Some pixels included in an objective area are assigned

Fig. 4. The flowchart of garment detection method.

as a foreground seed for the first detection. This area is decided by other method and provided to this process.

(3) Executing the Grabcut

At the first time, this process executes the Grabcut using the seed defined in (2). From the second time, this process executes the Grabcut using the seed defined in (6).

(4) The most reliable area extraction

After executing the Grabcut, objective areas $a \in \{a_0, \dots, a_n\}$ will appear. As described at top of the section, only one area is required. Therefore, this process calculates a reliability about each area according to $Z(\cdot)$. The $Z(\cdot)$ is provided to the process by other method. Then this process decides the most reliable area a^* .

$$a^* = \arg \max_a Z(\cdot) \quad (2)$$

(5) Condition branches

The first branch confirms the number of Grabcut iteration. If the number passes a constant, this process moves to (7). The constant is enough large value and typically it is 100.

The second branch confirms the size of the a^* . It defines a^{**} to be a^* of just previous times. If a score calculated according to the equation (3) passes a constant, this method moves to (7). The constant is typically 0.00001.

$$\text{score} = |1 - \frac{a^{**}}{a^*}| \quad (3)$$

(6) Assigning pixels as new seed

Some pixels included in the a^* are assigned as a new

Fig. 5. The flowchart of iterating the Grabcut using the most reliable seed.

seed. This seed is assumed to be more effective than previous seeds because it is created from reliable area regarding the objective. Therefore, this method is able to detect a more accurate area by using the seed.

(7) Outputting result image

After enough executing the iteration from (3) to (6), this process throws the execution result to other method.

That is all processes of the method.

3.2 Face detection

In the proposed method, it detects faces by using the HOG feature as first procedure. The HOG⁽⁷⁾ is shape based feature. It is often adopted to detect fixed shapes like a face or a vehicle. By the detection, this procedure obtains the location of the face $f \in \{f_0, \dots, f_n\}$. Then it clips areas $b \in \{b_0, \dots, b_n\}$ based on the face location according to

$$\begin{aligned} b_{x(\text{top})} &= f_{x(\text{top})} - f_{x(\text{height})} * 2 \\ b_{x(\text{left})} &= f_{x(\text{left})} - f_{x(\text{width})} * 2 \\ b_{x(\text{height})} &= f_{x(\text{height})} * 6 \\ b_{x(\text{width})} &= f_{x(\text{width})} * 5 \\ x &= \{0, \dots, N\} \end{aligned} \quad (4)$$

The range of $\mathbf{b}$ is defined to cover the personal upper-body. Furthermore, to reduce calculation cost, it executes down sampling to $\mathbf{b}$. By the sampling, both height and width of an image become half size.

3.3 Neck detection

The neck is assumed to locate under the face. Therefore, this procedure estimates an edge line of the neck by detecting the face area. For the estimation, it uses the iterating the Grabcut (Fig 5). It provides the f_x ($x \in \{0 \text{ to } n\}$) to the method as first seed area. Furthermore, it provides the $Z(a)$ according to

$$Z(a) = \text{pixels within } a \quad (5)$$

By assigning f_x as the first seed, the detected areas in 3.1(3) will be very small without an area including the f_x . Therefore, $Z(a)$ is simplified.

After executing the iterating the Grabcut, it obtains a face area. A neck area is defined as an area below the f_x . Furthermore, it executes corner finding to the neck area. By the process, corners are detected. Finally, the neck edge line is created as a group of corners $\mathbf{c} \in \{c_0, \dots, c_M\}$ by chaining the corners from the most left one to the most right one. If the over two lines exist, calculate a distance from the center of f_x to both corners. A line indicating the most little score is assumed as true line.

3.4 Area detection to create a seed

This procedure executes the iterating the Grabcut (Fig 5) to three areas $\mathbf{a}^{\text{seed}} \in \{a_0^{\text{seed}}, a_1^{\text{seed}}, a_2^{\text{seed}}\}$ around the neck line. These areas locate on the most left, the most right and center of the neck line. The height of areas are $f_{x(\text{height})}$ and the width of areas are $f_{x(\text{width})}/3$. Furthermore, it provides the $Z(a, \mathbf{a}, \mathbf{c})$ according to

The $Z(a, \mathbf{a}, \mathbf{c})$ indicates that the detected area a which is close to many corners $\mathbf{c}$. The a should be the a^* . Eventually, each a^{seed} is equal to each a^* .

After iterating the method to each area, it decide an area $a^{\text{seed}*}$ as a seed for the garment detection. It decided according to

$$a^{\text{seed}*} = \arg \max_a Z(a^{\text{seed}}, \mathbf{a}^{\text{seed}}, \mathbf{c}) \quad (7)$$

An objective of the equation (7) is same to the equation (6).

Furthermore, from the second (means next person) seed creation, it is able to obtain an effective $a^{\text{seed}*}$ by reducing any garment area from the a^* at any iteration.

3.5 Garment detection

As final procedure, it executes garment detection using

the $a^{\text{seed}*}$ as the first seed of the iterating the Grabcut (Fig

$$Z(a, \mathbf{a}, \mathbf{c}) = \sum_{i=0}^M \begin{cases} 1 & \text{if } dst(a) = \min(dst(\mathbf{a})) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$dst(a)$ = the shortest distance between a to c_i

5). The equation (6) is assigned to equation (2) to decide the most reliable area. After finishing the method, the a^* is defined as a garment area in the image.

Furthermore, from the second garment detection, it is also able to detect a garment efficiently by reducing other garment area from the $a^{\text{seed}*}$ at any iteration.

That is all processes of the proposed method.

4. Examination

To evaluate the proposed method, we conducted an examination of garment detection. As test data, we collected 155 images containing over one child. Their pose is various. 154 faces are detected from the data by the face detection procedure described in section 3.2. To compare the proposed method, "simplified garment detection method" and "garment detection method by voting" are executed in the same situation. As comparison criterion, f1 measure is adopted. This score is calculated to according to

$$f1 \text{ measure} = \frac{2 * precision * recall}{precision + recall}$$

$$precision = \frac{tp}{tp + fp} \quad (8)$$

$$recall = \frac{tp}{tp + fn}$$

The tp is ground truth corresponding to detected garment area. The fp is non ground truth corresponding to detected garment area. The fn is ground truth corresponding to not detected garment area. The Table 1 shows the experimental result. The precision is an average score of all precision of detection. The recall is an average score of all recall of detection. The f1 measure is an average score of f1 measure

Table 1. Experimental result.

	Proposed method	Simplified method	Voting method
Precision	0.646854	0.352029	0.178684
Recall	0.652573	0.460414	0.263935
F1 measure	0.606775	0.363399	0.188696

Fig. 6. Pixels in each face area and each F1 measure.

Fig. 7. F1 measures of each method.

of detection. From the f1 measures of Table 1, the proposed method is confirmed that it has good performance than other method.

From the experiment, we suspected that a first seed size for the Grabcut causes error of each score of Table 1. To confirm the suspicion, the Fig 6 shows a relationship between the pixel amount of each face area and each f1 measure. As the graph shows, the f1 measure do not relate on the face size. This means that the quality of the proposed method do not depend on image size. Therefore, some unknown problems which cause the error are existing in other procedures.

To compare behavior characteristics of the each method, the Fig 7 shows f1 measures of them. Their scores are arranged in order of size like the Fig 6. From the graph, it is assumed that their behaviors are similar because fluctuations of each f1 measure move similarly. Therefore it is assumed that any method contains same problems that cause the error of f1 score, so discovering problems of simplified method should contribute to improve the proposed method.

Finally Fig 8 shows some images that created as the experiment result. The left images are clipped image based on the face coordinates. The next images are ground truth images which were detected manually. The third images are

results of the proposed method. The fourth images are results of the simplified method. The final images are results of the voting method. The 5 images from top indicate that the proposed method performed correctly. On the other hand, the 2 images from bottom indicate that other method is better than the proposed method.

5. Conclusions

In this research, a garment detection method is proposed. This method iterates the Grabcut using the most reliable seed. From an examination, it is confirmed that it detects a back person's garment. On the other hand, the accuracy is not enough because some areas which is not detected have been existing yet. To improve the method, it is required to discover problems lurking in it. Furthermore, as one way, including the strength of the model based method to this method is possible, or re-defining the seed area is also important.

References

- (1) Kota Yamaguchi, Hadi Kiapour, Luis E Ortiz, Tamara L Berg : "Parsing clothing in fashion photographs", 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3570-3577, 2012
- (2) Eichner, M. and Marin-Jimenez, M. and Zisserman, A. and Ferrari, V : "Articulated Human Pose Estimation and Search in (Almost) Unconstrained Still Images", International Journal of Computer Vision, Vol 99, No. 2, pp. 190-214, 2012
- (3) Yi Yang, Deva Ramanan : "Articulated Pose Estimation with Flexible Mixtures of Parts.", IEEE Conference on Computer Vision and Pattern Recognition, pp 1385-1392, 2011
- (4) Nozomi matsuda, Lifeng Zhang : "Kindergarten children Identification Method Using Clothing Feature", The Seventh International Conference on Information, Pages, Publication 2015
- (5) C. Rother, A. Blake, and V. Kolmogorov : "GrabCut - interactive foreground extraction using iterated graph cuts", ACM Trans. on Graphics, Vol 23, No 3, pp 309-314, 2004
- (6) Jeong Mun Kim, Hayato Yamana : "Clothing Segmentation using Voting Region Images", DEIM Forum, 2012
- (7) Dalal. N, Triggs. B : "Histograms of Oriented Gradients for Human Detection", IEEE CVPR, pp 886-893, 2005

Fig. 8. The Results of the proposed method.