Unique Feature Extraction for Efficient Communication

Jin Sam KWAK, Ju Hyung SON

***Abstract*—This paper was originally submitted to Xinova as a response to a Request for Invention (RFI) focusing on new Decision and Response Techniques for Security Applications. This paper describes an efficient method of communication between control room and field agent using unique feature extraction.**

In more detail, this discloses how to select certain context-dependent features to facilitate describing suspects in a crowd.

I. Abstract

IT is very important to have efficient communication between the security agents dispatched to the site and the control room monitoring CCTVs. For instance, when a control room passes information about a suspect to field security agents, the information should specify unique features of the suspect that can be easily identified by the field security agents. Likewise, when a field security agent detects a suspicious person and passes information about the person to a control room, it is necessary to specify unique features of the suspicious person that can be easily identifiable on the CCTV monitor.

'Easily identifiable features' means that they should be easily recognizable from the point of view of the receiving side, not of the sending side. These features could be an object's own features, but in many cases, they would be something relatively distinguishing in relation to the surroundings. For example, if an adult whose height is 175 cm is standing among other adults with average height, the information of 175 cm tall, which is an object' s own feature, might not be distinguishable. On the other hand, if he is standing among children, his height which didn't provide meaningful discrimination in the previous circumstance can be an easily identifiable feature.

II. Summary of the Invention

In this SR, a new method is proposed for more efficient communication between a control room and field security agents when they inform detailed description of a suspect to each other. This method utilizes extracting unique features of a suspect which discriminate the suspect from the surroundings.

This document was submitted to Xinova on Feb 2, 2017.

Jin Sam Kwak and Ju Hyung Son are with WILUS Institute of Standards and Technology Inc., Seoul, Korea

In order to achieve this, it is proposed to extract unique features in the following manner:

- Extract features of a suspect (or target)
- Extract features of the people around the suspect
- Count the number of people per each feature extracted from the people around the suspect
- Among the features of the suspect, select the features with fewer counts (at this point, it is possible to give weighting to each feature depending on its level of discrimination)
- Provide the selected features to use the communication between a control room and filed security agents

III. Detailed Explanation

This SR proposes a new method to extract unique features which can be utilized to specify a suspect for more efficient communication between a control room and field security agents. As mentioned in Section 1, the unique features here could be an object's own features or could be very common features. In other words, the unique features are determined by relative comparison with surrounding objects and chosen for the control room or field security agents to easily identify with the naked eye.

For this purpose, this SR proposes to extract unique features through the following steps:

1) Point out/select the suspect (or the target)

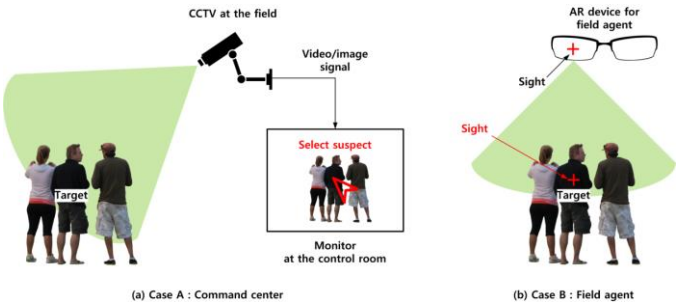


Figure 1. Point out/select the suspect

First of all, the person who wants to deliver information of the suspect needs to specify the suspect. Referring to Figure 1, the

agent in the control room may select the suspect by clicking the target on the display image (all part of the image will be analyzed through following steps). At this time, the image pixels related to the suspect may be separated from the background/surroundings.

If the proposed algorithm is processed at field agent's AR glasses, the field agent can point out the suspect by his/her finger (and the AR glasses will calculate the direction of the finger) or a sight of AR glasses. If is not, the control room selects candidates of the suspect by referring the location (and head direction) of the field agent and the video/image from CCTV located near the field agent.

2) Extract features of a suspect (or target)

Extract features of a target object through machine vision analysis. For instance, the following elements can be the features: gender, skin color, height, body type, hairstyle, whether or not wearing accessories (such as glasses, hats, bracelets, etc.), color of clothing, form of clothing, whether or not having supplementary object (such as bags), color of the supplementary object, form of the supplementary object, and so on.

The features utilized in the SR may be classified into two main features - a motion feature and an appearance feature.

Table 1. Feature Classification

Feature Class 1	Feature Class 2	Feature Class 3
Motion features	overall posture & movement	sitting down / standing / lying down / walking / running / …
	people arrangement (degree of dispersion)	in a group (number of neighboring people) / alone
	head direction	up / down / right / left
	arm position	…
	leg position	…
Appearance features	basic appearance	overall color arrangement - all black / black and white / …
		gender - male / female
		age group - child / adolescent / adult / …
		race
		hair style & length
		hair color
		body shape - thin / mid-range / fat / …
		height - small / mid-range / tall / …
	clothing	overall type - one-piece / top & bottom
		top - type / shape & color
		bottom - type / shape & color
		specific outfit / costume
	accessory	glasses
		hat
		gloves
	other objects	bag / …

In this SR the appearance features may be mainly utilized, but the motion features can also be utilized.

In addition, people who are very close to the suspect can also be included in the features. In such a case, the corresponding feature includes the physical distance between the suspect and the person close to the suspect. For instance, the following can be features: "the person on the very right side of the suspect is a male, white, 180cm tall, skinny, bald, wearing a blue jacket and having a briefcase in his left hand."

3) Extract features of people around suspect

Extract features of the people around the suspect in the same manner.

Herein, two of additional information - the head direction (or FOV) of the field agent and the distance between the suspect and the field agent can be referred while extracting features of the surrounding people.

Let's suppose the situation that the control room is informing the suspect to the field agent. The computer/server of the control room may extract features from the image/video transmitted from CCTV near the suspect.

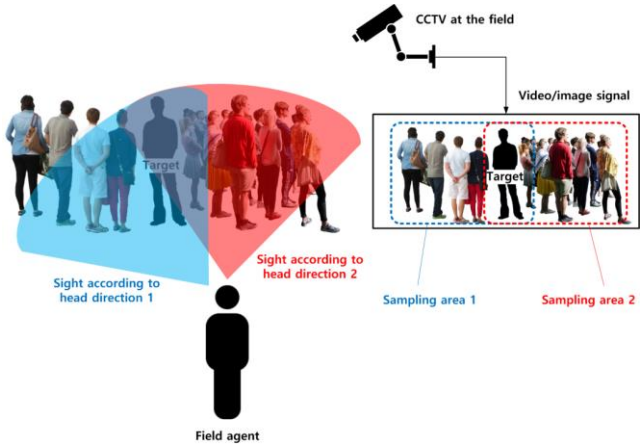


Figure 2. FOV of the field agent according to the head direction

According to the SR, the field agent can transfer the head direction information to the control room by using various sensors. Otherwise, the image of the field agent captured from CCTV can be analyzed, and as a result, the control room can aware the head direction information of him/her.

At this point, the field agent may see the different group of people around the suspect for the different head direction and the context of two groups (and surrounding environment) may differ to each other. As a result, features from different group may differ to each other, too. Therefore, the control room may set sampling area corresponding to the head direction of the field agent and extract the features therefrom. In Figure 2, the control room may utilize sampling area 1 when the field agent is heading to head direction 1 and utilize sampling area 2 when the field agent is heading to head direction 2.

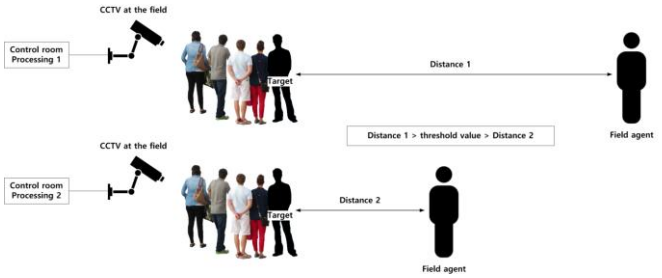


Figure 3. A distance between the field agent and the suspect

The control room can calculate a distance between the field agent and the suspect based on their locations. In general, it is hard for an observer to recognize a target object when the target object is really far from the observer. Likewise, when the field agent is far from the suspect, the field agent may not recognize the detailed feature information of the suspect. Therefore, if the distance between the field agent and the suspect is larger than a preset threshold value, the control room may extract features having high discrimination - such as a color, an overall motion - through certain processing method (algorithm) (the upper image of Figure 3. Besides, if the distance between the field agent and the suspect is smaller than the preset threshold value, the control room may extract detailed features through various processing methods (algorithms) (the lower image of Figure 3. At this time, the control room may extract more features than the former case.

By utilizing different processing methods according to the distance, the control room can process the features more efficiently, and the amount of information to be processed can be decreased.

4) Count the number of people per each feature extracted from the people around the suspect

Count the number of people who have the same feature extracting from the people around the suspect. According to the previous Step (3), the kinds and the numbers of features may be determined based on the head direction of the field agent and the distance between the field agent and the suspect. The following table shows an example of the result. Let's assume that the total number of people is 80. The sum of the number of people per each feature could be less than 80 or exceeding 80 because machine analysis might fail to analyze a certain feature (for example, not able to distinguish gender of a person) or some features (e.g., clothing) can be overlapped.

Table 2. The features extracted from the people around the suspect and the weights

Feature Class 1	Feature Class 2	Feature Class 3	Number of People	Weighting
Motion Features	overall posture & movement	running	34	1.30
		standing	23	1.10
	…	…	…	…
Appearance Features	gender	male	47	2.50
		female	28	2.50
	age group	child	8	1.30
		adolescent	20	1.20
		adult	25	1.00
		middle aged	14	1.00
		senior	1	1.20
	clothing	short pants	15	0.85
		long pants	10	0.85
		skirt	15	1.20
		jacket	18	1.52
		sportswear	22	1.45
	accessory	glasses	13	1.74
		hat	10	1.84
		gloves	8	0.65
		…	…	…

In the table, the weighting is a value reflecting the level of discrimination of the feature. Of course, these weighting would also differ depending on the context. For instance, when people are sitting in an auditorium watching a classical music concert the length of pants would not be an appropriate metric as discrimination parameter, so this feature is given low weighting value. On the other hand, when people are standing in a gallery watching a golf match, the length of pants as the level of discrimination is meaningful so this feature is given high weighting value. In addition, the level of discrimination would differ depending on the location of the object even in the same circumstance. For instance, in the case of a golf course gallery, it is obvious that the level of discrimination of the color of pants would differ depending on whether he/she is standing in the front row or in the back row.

According to the SR, kinds of the features to be counted may be determined based on the context of the surroundings of the suspect. Moreover, a value of each weighting may be assigned corresponding to the context, too.

For example, the control room can extract features from the image/video of the security area at certain time (or period) over an ordinary phase. When some of the features have been extracted more than predefined threshold, the control room can determine not to count the number of corresponding features because of a low discrimination level. Otherwise, the control room may assign a low weighting value to the corresponding features due to the same reason.

On the other hand, the respective feature may have a basic weighting value based on the nature of the human visual perception. For instance, the feature related to a relatively small object/accessory such as an earring may have a low basic weighting value, because the human eye barely recognizes the earring that several meters apart. In addition, the control room determines a correction value of each feature based on the context. Let's see the next figure.

Figure 4. FC Barcelona's jersey and the supporters in the stadium

Basically, the feature related to FC Barcelona's jersey may have a high basic weighting value (e.g., 4.0) according to its size and color contrast. However, the most of supporters in the stadium may put the same jersey on during the game and the level of discrimination of the feature related to the jersey become very low, resulting in a low correction value (e.g., -3.7). The final weighting value can be calculated based on the combination of the basic weighting value and the correction value. According to previous example, the feature related to the jersey in the stadium may have the final weighting value 0.3 (4.0 - 3.7).

Moreover, the weighting value can be determined based on the distance between the field agent and the suspect. Referring to the Figure 3, the field agent may be hard to recognize the suspect's detailed arm/foot position (motion features) when they are too far from each other. In this case, the motion features related to the arm/foot position may have a low weighting value. However, as the distance decreases, the weighting value of the motion feature related to the arm/foot position may increase.

5) Among the features of the suspect, select the features with fewer counts

In the next step, among the above features, count the number of people per each feature that applies to the suspect, and then select the features with fewer counts.

For instance, let's assume that the suspect is an adult female with long pants and wearing a hat, glasses and gloves. In this case, the number of people who have the same features as the suspect's among the entire features of the people around her is shown in the following table. Here, the item with the lowest count is gloves, followed by hat and glasses. On the other hand, the features of adult and female are shared with a relatively large number of people so the levels of discrimination are low. Therefore, it would be effective to say, "The target is wearing glasses, a hat, and gloves." to easily identify the suspect.

Table 3. The weighted counts of the features

Feature		Number of people	Weighting	Weighted Count
ender	female	28	2.50	28*1/2.5=11.2
age group	adult	25	1.00	25*1/1.0=25
clothing	long pants	15	0.85	15*1/0.85=17.65
accessory	glasses	13	1.74	13*1/1.74=7.47
	hat	10	1.84	10*1/1.84=5.43
	gloves	8	0.65	8*1/0.65=12.31

As described earlier, it is possible to consider the level of discrimination of each feature in extracting unique features. Taking this further into account, gloves might have a low level of discrimination and gender has a high level of discrimination. In the table above, the weighting is given by calculating

weighted count = count * (1/weighting)

as an example. This allows us to select some features - there can be referred to as 'dominant features' - with lowest weighted counts and extract 'a female wearing glasses and a hat' as unique features. (For an effective communication, the features can be transmitted in decreasing order of weighted counts (e.g., transmitting the messages in order of 'hat', 'glasses' and 'female'.))

Now, we can consider a case that there are pluralities of people having the dominant features.

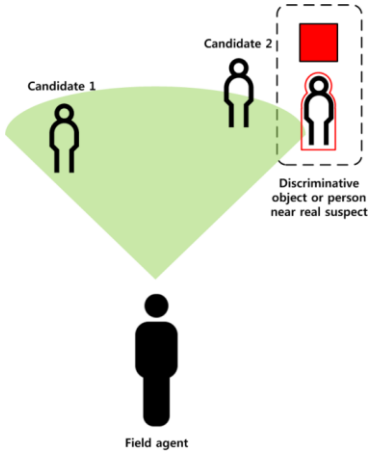


Figure 5. Pluralities of people having the dominant features

In this case, the control room may deliver additional information related to the direction of the suspect when the control room knows the FOV of the field agent. Referring to the Figure 5, when the real suspect is Candidate 1, the control room may deliver the dominant features with the direction information 'left of the FOV' to the field agent. Besides, the control room can transmit additional information related to the discriminative object or person near the real suspect. When the real suspect is Candidate 2 in the Figure 5, the control room may deliver the features of the discriminative object or person and may say 'find the suspect having the dominant features xxx around the discriminative object X'.

Besides, the control room can easily determine the number of the dominant features to be transmitted. First, the weighted count value of the features can be listed in decreasing order. Referring to the Table 3, the order of the features can be listed as follow - hat, glasses, female, gloves, long pants and adult. For an efficient communication, the number of dominant features have to be transmitted can be determined based on the number of people having combination of dominant features. According to the Table 3, the number of people having 'hat' are 10. Herein, let's assume the number of people having each combination of the dominant features as following table.

Table 4. Number of people according to the combination of dominant features

Combination of dominant features	# of people
'hat'	10
'hat' and 'glasses'	1
'hat' and 'glasses' and 'female'	1
'hat' and 'glasses' and 'female' and 'gloves'	1
'hat' and 'glasses' and 'female' and 'gloves' and 'long pants'	1
'hat' and 'glasses' and 'female' and 'gloves' and 'long pants' and 'adult'	1

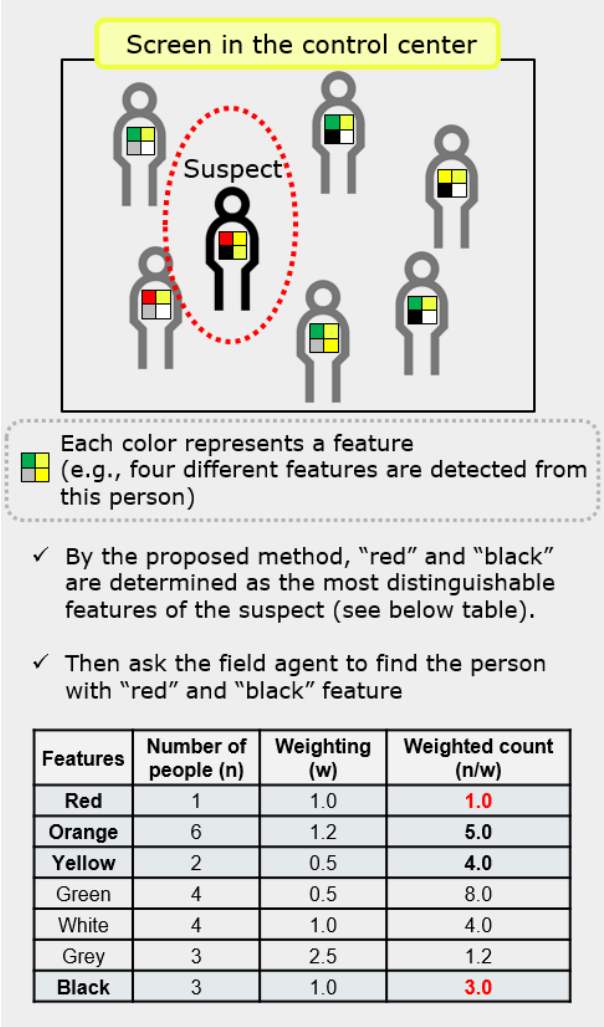
In the Table 4, there are only one (or a certain predetermined number) person who has two features - 'hat' and 'glasses'. Therefore, the control room doesn't have to transmit additional dominant features such as 'female', 'gloves' and the like.

In the above description, it has been explained how to independently figure out and count each feature for a brief description. However, it is possible that one feature is composed of several sub-features. For example, there might be 20 female and 20 people wearing hats, but the number of female wearing a hat can be 1. In this case, both the gender and wearing a hat cannot be features if each feature is considered independently. However, if 'female wearing a hat' is considered as a feature, it is possible to extract a unique feature much more effectively. Furthermore, in order to reduce computational load in step (2) and (3), it is also possible to directly/quickly extract the feature information of the people around the suspect based on the feature information set of the suspect (e.g., female adult, long pants, wearing a hat/glasses/gloves) acquired in step (2). In this case, step (3) and (4) can be performed selectively so it is possible to quickly obtain the resulting value of step (5).

Application Scenario:

- The application scenario in which this proposed method can be used in a control room is as follows:
  i. When a suspect is detected on a CCTV monitor, the operator in the control room selects the suspect by clicking on the location of the suspect.
  ii. The system extracts unique features of the selected suspect according to the proposed method, and then instructs them to the operator.
  iii. The operator passes the feature information to the field security agents using radio communications and such.

- The application scenario in which this proposed method can be used by field security agents is as follows:
  i. When a suspect is detected at the site, a field security agent informs the system of the location of the suspect through certain action, such as simply glaring at the suspect. Since it is not easy for the field security agent to clearly select the suspect, the system extracts features of the candidates that the field security agent seems to have pointed out.
  ii. Afterwards, when the system informs the field security agent of the extracted features, the field security agent selects the feature applying to the suspect he has pointed out among the extracted features.
  iii. The field security agent passes the information to the control room.



Features	Number of people (n)	Weighting (w)	Weighted count (n/w)
Red	1	1.0	**1.0**
Orange	6	1.2	**5.0**
Yellow	2	0.5	**4.0**
Green	4	0.5	8.0
White	4	1.0	4.0
Grey	3	2.5	1.2
Black	3	1.0	**3.0**

Figure 6. Concrete example

IV. Conclusion

This method brings with it several merits:

- Selected features are "context dependent" unique features
- Easier to distinguish the suspect in verbal communication

Acknowledgment

This research was originally submitted to Xinova, LLC by the author in response to a Request for Invention. It is among several submissions that Xinova has chosen to make available to the wider community. The author wishes to thank Xinova, LLC for their funding support of this research.

More information about Xinova, LLC is available at www.xinova.com.