ANALYSIS OF APPEARANCE FEATURES FOR HUMAN MATCHING BETWEEN DIFFERENT FIELDS OF VIEW

Xiaowei Lv¹, Qing-Jie Kong¹,², Fei Weng¹, Yuncai Liu¹

¹Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong Univ., Shanghai, China
²Beckman Institute, Department of Electrical and Computer Engineering, Univ. of Illinois at Urbana-Champaign, Urbana, IL, USA

ABSTRACT

Human matching between different fields of view is a difficult problem in intelligent video surveillance; whereas fusing multiple features has become a strong tool to solve it. In order to guide the fusion scheme, it is necessary to evaluate the matching performance of these features. In this paper, four typical features are chosen for the evaluation. They are the Color Histogram, UV Chromaticity, Major Color Spectrum Histogram, and Scale-Invariant Features (SIFT). Quantities of video data are collected to test their general accuracy, robustness, and real-time applicability. The robustness is measured under the conditions of illumination changes, Gaussian and salt noises, foreground errors, resolution changes, and camera angle differences. The experimental results show that the four features bear distinctive performances under the different conditions, which will provide important references for the feature fusion methods.

Index Terms— Performance Evaluation, Appearance Features, Human Matching, Different Fields of Views, Video Surveillance

1. INTRODUCTION

In video surveillance, the field of view (FOV) of a camera is limited; furthermore, it is difficult to cover all surveillance regions with a great deal of overlapped cameras due to the enormous costs and the computing complexity. Therefore, people have to implement surveillance with non-overlapped cameras. In this context, video surveillance in non-overlapped cameras becomes a hot topic in recent years. This field associates with a series of research directions, including the human matching, topology estimation, data association etc. Among them, human matching between different FOVs is a foundational work.

In current research relating human matching, varieties of appearance features have been chosen to build the appearance model. For instance, the Color Histogram (CH), UV Chromaticity (UVC), Major Color Spectrum Histogram (MCSH), and Scale-Invariant Features (SIFT). However, due to the great impacts on multiple factors (e.g., illumination, detection errors, camera parameters) between different cameras, only single feature can not achieve high matching accuracy. In this case, some researchers attempted to fuse two or more features to match objects and attained higher accuracy. For example, Patwardhan et al. integrated the CH and the SIFT to match human [1]; Madden and Piccardi built a framework to fuse height of persons and the MCSH [2]. However, they usually fused multiple features only by a simple framework such as Bayesian framework, Gaussian framework, without considering difference involved in detecting conditions. Therefore, if knowing matching performance of these features under various detection conditions, we can find which features should be use, and how to use them for human appearance matching.

In this paper, we experimentally evaluate the property of the four representative appearance features in human matching, which are the CH, the UVC, the MCSH, and the SIFT. Because the three kinds of color features can not contain any spatial information, we partition the foreground

This work is supported in part by China National 973 Program of 2006CB303103, China NSFC Key Program of 60833009 and China National 863 Project of 2009AA01Z330.
of a human into three parts: head, torso and legs, using the method in [7], and match them respectively.

3. EVALUATION METHODOLOGY

In this section, we propose the evaluation measurements in the five kinds of interferences.

3.1 Illumination variation

A large proportion of matching errors comes from the illumination difference between different FOVs. In order to represent the illumination changes in measuring the algorithms’ robustness, an intensity descriptor is presented as follows:

\[ I = \left[ 1 - \frac{E(v_a)}{E(v_b)} \right] \times 100 \]  

where \( v_a \) and \( v_b \) are the intensity of each pixel in Camera A and B, respectively, herein we let \( v_b \geq v_a \); function \( E(\cdot) \) means to compute the expectation.

3.2 Detection noises

In testing of noise robustness, Gaussian and Salt noises are increasingly added to the foregrounds of the ground truth with a range between 0 and 80%.

3.3 Foreground errors

Due to the difficulty in foreground segmentation of videos, the detection errors of human foregrounds are another kind of common interferences to the human matching. We define the accuracy descriptor to denote the foreground errors, introduced as follows.

Let \( S \) be the perfect foreground, with no background pixels in \( S \), and all of the foreground pixels are in \( S \). Let \( S' \) represent the foreground with both holes and redundancy, and \( S_f \) denote the \( S' \) filled holes. Then, a parameter \( P_1 \) is defined as \( P_1 = \frac{S \cap S'}{S} \), in which \( \cap \) means to compute the intersection. Another parameter is defined as \( P_2 = \frac{S \cap S_f}{S_f} \). Thus, we define the foreground errors as:

\[ e = 1 - \frac{1}{2n} \sum_{i=1}^{n} (P_1_i + P_2_i) \]  

where \( 0 \leq e \leq 1 \); 0 indicates a perfect foreground, and 1 means that there are no real foreground pixels in the detected foreground. \( n \) denotes the number of the human in the dataset.

3.4 Resolution changes

The resolution of videos directly impacts the matching performance of the algorithms. Generally, the higher resolution is, the more detailed information the object’s appearance contains. However, by experiments we found that the matching accuracy does not obviously increase when the resolution is higher than a threshold. Therefore, we focus on the resolution changes under the threshold, and test the algorithms in the lower resolution. We define a resolution factor to represent the normalized variances between the experimental foregrounds and the original image. 0 indicates no changes and 1 indicates the size of the foreground is 0.

3.5 Camera angle differences

Shooting angle differences between two cameras also build a huge obstruction against the human matching. In this work, camera angle differences are changed in the range between 0 and 90 degree.

4. RESULTS AND DISCUSSIONS

In our experiments, no appearance transfer model is used to remedy the condition changes, so as to reflect the features’ essential characters. The dataset consists of 16 groups of ground truth pairs for the general accuracy. The 16 groups of foreground data are synthesized in every sort of the interferences. Each group of them has 10 different person pairs. Figure 1 shows a part of our ground truth dataset. Each pair of the foregrounds was shot by two non-overlapped cameras in different scenes.

4.1 General accuracy

Figure 1. A part of ground truth dataset.
In this experiment, we tested the general accuracy of the four algorithms in the first dataset. The dataset included all of the interferences presented in Section 3. The accuracy was measured on three ranks, shown in Table I. The first rank means the object with top 1 similarity is the correct matching; the second rank means the objects with top 2 similarities contain the correct matching; and so forth. Also, the features are fusion in a Bayesian framework.

From Table I, we find the UVC algorithm shows the highest accuracy, while the CH algorithm has the lowest accuracy. Furthermore, it is also found that considering the objects with top 3 similarities can attain higher accuracy. It implies that we could achieve better effectiveness by fusing the information of the objects with top 3 similarities to make the final decision.

### 4.2 Robustness

The robustness was tested according to the measurements defined in Section 3. The illumination descriptor varied from 5% to 50%; the Gaussian and Salt noises were increasingly added from 0 to 80%; the foreground errors changed from 2% to 38%; the resolution factor varied between 0 and 0.7; and the camera angle differences are regulated between 0 and 90 degree. When one factor was tested, the other factors that might affect the accuracy were set nearly zero or non-changeable. The experimental results are illustrated in Figure 2.

Figure 2(a) shows the illumination robustness. From this figure, we find the accuracy of both the UVC and SIFT algorithms can hardly be affected by the illumination changes, whereas the accuracy of the other two algorithms rapidly decreases with the illumination changes, because that the former two kinds of features contain little intensity component.

Figure 2(b) and (c) show the noise robustness. We find the SIFT based algorithm is more sensitive to the two noises than the other algorithms. The three color-related algorithms can retain their accuracy until around 60% Gaussian and Salt noises. Among them, the UVC algorithm shows the highest accuracy and the strongest robustness. The reason why the SIFT feature has the weakest robustness against the noises lies in that the SIFT feature is based on the texture details, and the increasing noises greatly disturb the extraction of the SIFT feature.

Figure 2(d) shows that the accuracies of the four algorithms all gradually decline with the increasing errors at the similar speed. This phenomenon reflects the fact that no features can keep well performance when the information of human foregrounds is heavily lost.

For the resolution changes, the UVC and CH algorithms embody much stronger robustness than the other two features, as is shown in Figure 2(e). In terms of camera angle differences, the three kinds of color based features show far stronger robustness than the SIFT feature. This implies that the SIFT feature might not be applied into the feature fusion in the case of huge shooting angle differences. The UVC still shows the best performance under the two kinds of interferences, respectively.

### 4.3 Real-Time costs

The evaluation of real-time applicability consists of two stages. The first includes the feature extracting and the appearance model building; the second includes the comparison of the similarity and the output of the matching results. In this experiment, we considered two kinds of resolution: the original resolution and the resolution factor of 0.545, which is a common situation and easy to attain. The experimental results are shown in Table II.

From the table, we can see that the time costs of all of the algorithms apparently decrease when the resolution declines. The time cost of the SIFT algorithm is much larger than that of the others. Therefore, it might not adapt to the real-time application without more improvements. The Color Histogram algorithm costs the least time to the original foregrounds, whereas the UV Chromaticity algorithm performs best when the resolution declines.
5. CONCLUSIONS

In this work, we experimentally compare and evaluate the human matching performances of the four major appearance features (the Color Histogram, UV chromaticity, Major Color Spectrum Histogram, and Scale-Invariant Features) on the impacts of the five kinds of common interferences. The general accuracy, robustness and real-time applicability are all measured. From the experimental results, we can draw the following conclusions:

1) The matching accuracy of every feature is below 80%. Therefore, it is necessary to fuse multiple kinds of features to acquire higher accuracy.

2) Better effectiveness may be obtained, if we can fuse the information of the objects with top 3 similarities to make the final decision.

3) Among the three color based features, UV Chromaticity feature shows the highest accuracy and strongest robustness. Therefore, it should be chosen to match human fused with other noncolor features.

4) The SIFT feature might not be fused, when shooting angle differences become huge. Also the time cost of the SIFT algorithm is too much to be used in real time, without more improvements.

In the future work, we will continue to evaluate the performances of more available features. Meanwhile, according to the results of the performance evaluation, we are going to fuse the efficient features to build the appearance model with higher accuracy, robustness, and real-time applicability.

6. REFERENCES


