

# ROBUST HUMAN TRACKING METHOD BASED ON APPEARANCE AND GEOMETRICAL FEATURES IN NON- OVERLAPPING VIEWS

Binbin Liang<sup>1</sup>, Songchen Han<sup>1</sup>, Yan Zhu<sup>2</sup>, Liping Di<sup>1</sup>

<sup>1</sup>School of Aeronautics and Astronautics, Sichuan University, Chengdu, China  
liangbinbin110@126.com

<sup>2</sup>College of Civil Aviation, Nanjing University of Aeronautics and Astronautics,  
Nanjing, China  
735506286@qq.com

## **ABSTRACT**

*This paper proposes a robust tracking method which concatenates appearance and geometrical features to re-identify human in non-overlapping views. A uniformly-partitioning method is proposed to extract local HSV(Hue, Saturation, Value) color features in upper and lower portion of clothing. Then adaptive principal view selecting algorithm is presented to locate principal view which contains maximum appearance feature dimensions captured from different visual angles. For each appearance feature dimension in principal view, all its inner frames get involved in training a support vector machine (SVM). In matching process, human candidate filtering is first operated with an integrated geometrical feature which connects height estimate with gait feature. The appearance features of the remaining human candidates are later tested by SVMs to determine the object's existence in new cameras. Experimental results show the feasibility and effectiveness of this proposal and demonstrate the real-time in appearance feature extraction and robustness to illumination and visual angle change.*

## **KEYWORDS**

*Human Tracking, Non-Overlapping Views, HSV Appearance, Geometrical Features, SVM*

## **1. INTRODUCTION**

Video-surveillance is an increasingly developed technology in various domains such as on-site security surveillance, shoplifting evidencing and intelligent recognition. In many circumstances, video-surveillance system consists of multiple cameras without overlapping views, thus making spatial and temporal information of moving objects unavailable for parts of the path. So it's significantly desirable to explore a robust and real-time human tracking method in non-overlapping views.

The human tracking task is usually composed of three main aspects, namely, foreground detection, feature extraction and object matching. Foreground detection is a fundamental David C. Wyld et al. (Eds) : NETCOM, NCS, WiMoNe, CSEIT, SPM - 2015  
pp. 219–235, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51619

component in visual tracking system. It extracts the people of interests and separates it with backgrounds. The performance of foreground detection directly determines the accuracy of human tracking. Feature extraction is a crucial component in visual tracking system. Human feature in computer vision consists of many descriptors which will help machine to understand the real world. Object matching is another crucial component in visual tracking system. When human features are extracted in independent surveillance area, they need to be matched from one area to another, making decision on “what is what” and “who is who”. The human tracking task in non-overlapping views depends upon these three aspects, especially the latter two aspects.

### 1.1. Related Work

While numerous acceptable techniques have been proposed so far for foreground detection, there is still a need to produce more efficient algorithms in term of adaptability to multiple environments, noise resilience, and computation efficiency. Related literatures suggest that Visual Background extraction (ViBe) outperforms recent and proven state-of-the-art methods in terms of both computation speed and detection rate<sup>[1]</sup>.

As for feature extraction, previous literatures have concentrated on appearance features to track human objects across disjoint cameras<sup>[2-6]</sup>. Gianfranco et al.<sup>[4]</sup> reviewed the development of a set of models that capture the overall appearance of an individual. Some of the models provide a holistic description of a person, and some others require an intermediate step where specific body parts need to be identified. Hyun-Uk et al.<sup>[5]</sup> segmented each person by a criterion with appearance and estimated the segmented regions as Gaussian mixture model (GMM) for correspondence. Both experimental results demonstrated that appearance feature has superior performance in identifying individuals, but illumination changes pose a major problem to the invariance of appearance features. In order to solve this problem, Javed et al.<sup>[6]</sup> proposed to compensate illumination variations using brightness transfer functions but his approach is limited to compensate for different illumination in different regions of the images. It still needs more researches on the solution of illumination variations in disjoint human tracking task.

Individuals can always be discriminated based on their appearances, except the notable situation where exists people in uniforms. In addition, appearance features will probably alter along with the change of visual angle which occurs frequently when people walk across cameras. Related researches combined geometrical features with appearance features and managed to overcome such problems to some extent<sup>[7-11]</sup>. Madden et al.<sup>[7]</sup> focused on a framework based on robust shape and appearance features extraction and performed well in some specific scenarios. However, his proposal only employed height as robust shape feature without considering the limitation of height in discriminating human beings due to the close resemblance of human stature. The popular remedy of this limitation is to combine gait feature with height to strengthen the discrimination of human beings. Takayuki et al.<sup>[9]</sup> proposed a method that tracks a walking human using the features of gait to calculate a robust motion signature and showed the potential validity of the proposed method in a typical surveillance system. Despite of the well-done performance of gait feature in human tracking task, it has deficiencies that a high identification rate and low computational cost are still far from being achieved<sup>[10]</sup>. Specifically, the typical gait feature is tough to adjust to the visual angle change in disjoint tracks.

In general, a discriminative appearance feature which is robust to illumination change and uniform disturbance needs to be extracted. Moreover, a discriminative geometrical feature which can adjust to the visual angle change expects more explorations.

## 1.2. Contributions of This Paper

This paper proposes a new robust human tracking method based on appearance and geometrical features in non-overlapping views. ViBe is initially adopted to detect and segment foregrounds of each single individual. Then a uniformly-partitioning method is presented to obtain the local low-bins HSV histograms. The uniformly-partitioning method is intended to eliminate the disturbance of illumination change. Thereafter, a new geometrical feature is created which integrates height estimate, height estimate difference parameters and the sinusoidal periodicity of walking. These three features can reflect the height and gait movement simultaneously. To our knowledge, it's original to integrate such three intrinsic features in human tracking which avoids the complexity of feature extraction and can robustly counter the visual angle change even across disjoint cameras. For a high hit rate, an adaptive principal view selecting algorithm (APVSA) and appearance feature dimension determining algorithm (AFDDA) are innovatively presented and SVM is also adopted for the same purpose. APVSA facilitates the robustness to appearance change and SVM strengthens the discrimination of human individual.

The proposed method has a configuration as shown in Fig.1. In this method, the foregoing cameras train and update SVMs. When new camera comes, all its inner frames are preprocessed and the foregrounds are detected. Then the appearance and geometrical features are extracted. The geometrical features filter the object candidates and the appearance features determine the tracked object continuously on the basis of principal view selecting mechanism. Section 2 introduces ViBe-based motion detection briefly. Section 3 and Section 4 describe appearance and geometrical features. Section 5 constructs the matching mechanism based on human features, APVSA, AFDDA and SVM. Section 6 discusses the experimental results in simple and complex scenario. Finally, Section 7 concludes the highlights of this proposal.

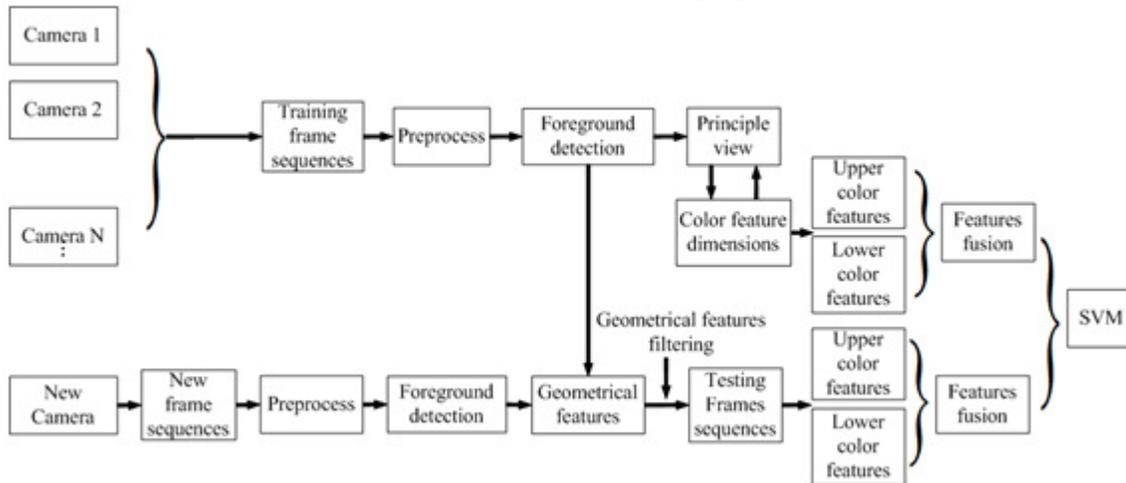


Fig.1. The process diagram for the proposed method

## 2. MOTION DETECTION

Motion detection is the key low-level fundamental work in intelligent video tracking task which requests accurate segmentation of foreground from background. The emerging ViBe algorithm is proven to perform satisfactorily in terms of both computation speed and detection rate. It provides an excellent robustness to scene changes and extracts foreground rapidly from early

frames. Thus, this paper employs ViBe to detect and segment moving objects in video scenes in the way of [12]. After an individual is segmented, an external bounding box is used to contain it.

### 3. APPEARANCE FEATURES EXTRACTION

This proposal is inspired by [7] to extract local appearance of upper and lower portion in clothing which can almost discriminate individuals instead of a global appearance. These features allow a more sensitive reaction to appearance change and fasten the extraction speed as well. In this paper, the upper 30-50 percent from the top of the external bounding box is chosen to be the upper portion, while the lower 65-85 percent from the top is chosen to be the lower portion.

Appearance features often vary with the illumination change. In order to deal with this, HSV color space is employed to model appearance color in this paper. HSV is a color space in which the effect from color brightness can be suppressed by decreasing value characteristic (V). In light of this, it can reduce the appearance dissimilarity caused by illumination change. Appearance feature is partitioned into upper and lower portion, and each of them will generate HSV color features. Moreover, each HSV component will generate a histogram and thus make massive bins which will cost a lot of time consumption. To cope with this, a uniformly-partitioning method of HSV space is presented on the basis of [13]. In this method, Hue component is divided uniformly into  $Q_H$  intervals, Saturation is divided uniformly into  $Q_S$  intervals and Value is divided uniformly into  $Q_V$  intervals. Consequently,  $H, S, V$  component is converted to quantization level as  $H_c, S_c, V_c$ , such that

$$H_c = \left\lfloor \frac{H * Q_H}{360} \right\rfloor, S_c = \lfloor S * Q_S \rfloor, V_c = \lfloor V * Q_V \rfloor, \quad (1)$$

Then  $H_c, S_c$  and  $V_c$  are integrated as a vector  $\gamma_{HSV}$  with different weight coefficients in the following formula

$$\gamma_{HSV} = H_c * Q_S * Q_V + S_c * Q_V + V_c, \quad (2)$$

This method reduces the amount of total histogram bins and improves the efficiency of appearance feature extraction. Moreover, since  $Q_S$  and  $Q_V$  are greater than 1, the weight of Value characteristic is lessened. This is conducive to the robustness to brightness change.

### 4. GEOMETRICAL FEATURES EXTRACTION

Height estimate is a commonly utilized geometrical feature because it keeps almost invariable across camera views. But height estimate is quite limited to identify individual because of the significant resemblance of people's height. The popular remedy tends to combine gait feature with height estimate to enhance individual's uniqueness. Nevertheless, the complication of gait features and poor robustness to visual angle change unveil its flaws in disjoint tracking tasks. In this paper, a new geometrical feature which is enlightened by gait movement is innovated.

#### 4.1. Geometrical Features

This paper assumes people in scenes nearly walk in a constant speed and stay in a same behavior model. In each frame, the apparent height of individual is defined as the vertical length from the

top of head to the bottom position of feet. Assuming Section 1 achieves accurate foreground segmentation and human being stays upright in most cases, the apparent height in this paper is computed as the length from the middle of the top pixel row to the middle of the bottom of bounding box instead.

The point  $p(x, y)$  on the image plane can be described with a homogeneous coordinate  $p = [x, y, 1]^T$  and its corresponding point in the world coordinate system is  $p_{real} = [x, y, z, 1]^T$ . The projection from the real world point  $p_{real}$  to the image point  $p$  is given by

$$\delta p = K[Rt]p_{real}, \quad (3)$$

where  $\delta$  is a nonzero scale factor and  $R, t$  denote rotation matrix and translation vector respectively.  $K$  is the intrinsic parameter matrix such that

$$K = \begin{bmatrix} f_u & s & u_0 \\ & f_v & v_0 \\ & & 1 \end{bmatrix}, \quad (4)$$

where  $f_u$  and  $f_v$  are the camera focal length expressed in pixel units along two axes,  $s$  denotes a skew coefficient.  $u_0, v_0$  represent the principal axis coordinate respectively.

Morphological researches indicate that human walking involves rhythmic up-and-down displacement of the upper body, leading to the apparent bobbing of head. Furthermore, these vertical movements must occur in a smooth sinusoidal manner for the conservation of energy<sup>14</sup>. The apparent height is exhibited as a sinusoidal curve  $h(t) = \mu_h + \sigma_h \sin(\omega t + \phi)$ . When the legs combine closest together, the maximum apparent height  $h_{max}$  occurs. When the legs separate furthest, the minimum apparent height  $h_{min}$  occurs.

Setting the maximum apparent height as the apparent stature, so the true height  $h_{stature}$  can be projected as below

$$h_{stature} = \delta [Rt]^{-1} K^{-1} h_{max}, \quad (5)$$

Let  $\Delta h_{i,i+1}$  denotes the apparent height difference between adjacent frames  $i$  and  $i+1$ . The mean  $\mu_\Delta$  and standard variance  $\sigma_\Delta$  of  $\Delta h_{i,i+1}$  in  $N_{biom}$  frames could be calculated statistically as

$$\mu_\Delta = \frac{\sum_{i=1}^{N_{biom}} \Delta h_{i,i+1}}{N_{biom}}, \quad (6)$$

$$\sigma_\Delta = \sqrt{\frac{1}{N_{biom}} (\Delta h_{i,i+1} - \mu_\Delta)^2}, \quad (7)$$

Another intrinsic feature, the walking periodicity  $F_{periodic}$ , equals to the frames number  $N_{periodic}$  counting from the previous  $h_{max}$  frame to the next  $h_{max}$  frame, that is

$$F_{periodic} = N_{periodic}, \quad (8)$$

Thus, three intrinsic features: height estimate  $h_{stature}$ , height estimate difference parameters  $(\mu_{\Delta}, v_{\Delta})$  and walking periodicity  $F_{periodic}$  are ascertained to be geometrical features. Notably,  $\mu_{\Delta}$  and  $v_{\Delta}$  depend on the vertical movements and  $F_{periodic}$  relies upon time interval, all of which avoid the complexity of computing in previous literatures and can keep unchanged in various visual angles.

## 4.2. Geometrical Features Based Matching

Whether two geometries match or not is determined by the similarity coefficient  $\rho_{biom}$  which is computed as below

$$\rho_{biom} = \alpha_{stature} norm_{stature} + \alpha_{\Delta} norm_{\Delta} + \alpha_F norm_{periodic}, \quad (9)$$

Here,  $\alpha_{stature}, \alpha_{\Delta}, \alpha_F$  represent the weighting coefficient of stature(height estimate), height estimate difference parameters and walking periodicity respectively such that  $\alpha_{stature} + \alpha_{\Delta} + \alpha_F = 1$ .  $norm_{stature}, norm_{\Delta}, norm_{periodic}$  are normalized functions in formulation (10), (11) and (12)

$$norm_{stature} = c_{stature} abs(h_{stature} - h'_{stature}), \quad (10)$$

$$norm_{\Delta} = c_u \lambda_{\mu} abs(\mu_{\Delta} - \mu'_{\Delta}) + c_v \lambda_v abs(v_{\Delta} - v'_{\Delta}), \quad (11)$$

$$norm_{periodic} = c_{periodic} abs(F_{periodic} - F'_{periodic}), \quad (12)$$

where  $c_{stature}, c_{\mu}, c_v, c_{periodic}$  denote normalization coefficients and they are valued with relevance to real need.  $\lambda_{\mu}, \lambda_v$  denote the weighting coefficient of  $\mu_{\Delta}, v_{\Delta}$  such that  $\lambda_{\mu} + \lambda_v = 1$ .  $abs(*)$  represents an absolute function.

## 5. OBJECT MATCHING

In this section, an adaptive principal view selecting algorithm (APVSA) will be introduced and the appearance feature dimensions in principal view are to be determined. For each appearance feature dimension, SVM is trained and is expected to fuse the upper and lower appearance feature together to generate a classifier. In matching process, below-threshold candidates will be filtered by geometrical features. Then the remaining candidates will be tested by the trained SVMs.

### 5.1. Principal View

Principal View (PV) refers to the view field which contains maximum appearance features captured from different angles, such as front, side and back. In non-overlapping wide area, principal view  $V_{principle}$  is selected manually and empirically. But the known image sequences are usually captured from non-principal views. In this case, an interim principal view (IPV)  $V_{interim}$  would be determined from the known views. This paper establishes an adaptive principal view selecting algorithm (APVSA) as presented in Algorithm 1.

**Algorithm1** Adaptive principal view selecting algorithm

---

```

Initial view  $V_{initial}$ 
if initial view is principal view  $V_{principle}$  then
    principal view is selected,  $V_{principle} = V_{initial}$ 
end if
else
    Set initial view as interim principal view,  $V_{interim} = V_{initial}$ 
    for each subsequent view  $V_i$  do
        if  $V_i$  is principal view  $V_{principle}$  then
            principal view is selected,  $V_{principle} = V_i$  break;
        end if
        else if the tracked object is contained in  $V_i$  then
            Count the appearance feature dimension number  $N_{D_i}$  in  $V_i$ 
            if  $N_{D_i} > N_{D_{interim}}$  then
                Substitute  $V_i$  for  $V_{interim}$ ,  $V_{interim} = V_i$ 
            end if
        end else if
    end for
end else

```

---

**5.2. Appearance Feature Dimensions**

In this paper, each over-threshold appearance feature is recognized as one appearance feature dimension (AFD). Different features in different angles have a unique AFD. Principal view is expected to have maximum AFDs.

Given there are  $N_{training}$  frames captured from principal view, their HSV histograms are generated and analyzed. Two histograms will be gathered into the same group if their similarity is larger than a set threshold. Otherwise, they will be categorized into two different groups. If the frames in a group outnumber the set threshold, this very group will generate an appearance feature dimension.

The similarity of two appearance features  $F_1, F_2$  is codetermined by the correlations of their upper appearances and lower appearances. The upper appearance generally contributes more to discriminate individuals, so it is weighted 65%. The lower appearance contributes less, so it is weighted 35%. Herewith, since the left half of HSV color histogram represents the upper appearance and the right half represents the lower appearance, the similarity of two histograms  $H_1, H_2$  is computed from the correlations of their left halves and right halves. The correlation of two histograms  $H_1, H_2$  is calculated statistically as

$$\rho(H_1, H_2) = \frac{\sum_{i=0}^n H_1(i)H_2(i)}{\sqrt{\sum_{i=1}^n H_1^2(i)H_2^2(i)}}, \quad (13)$$

where,

$$H'_k(i) = v_k(i) - \left( \sum_{j=1}^n v_k(j) \right) / n, \quad (14)$$

where  $v_k(i)$  denotes the value of dot  $i$  in histogram  $k$ . The similarity of two histograms is formulated as following

$$\rho_{app}(F_1, F_2) = 0.65\rho(H_{1,u}, H_{2,u}) + 0.35\rho(H_{1,l}, H_{2,l}), \quad (15)$$

Algorithm 2 shows the pseudo code of appearance feature dimension determining algorithm (AFDDA).

**Algorithm 2** Appearance feature dimensions determining algorithm

---

```

First frame sequence do
  Extract the appearance feature  $F_{initial}$ , set it as the first dimension,  $D_{initial} = F_{initial}$ 
  for Each subsequent frame  $frame_i$  do
    Extract its feature  $F_i$ 
    for Each feature dimension  $D_j$  in dimension set  $D$  do
       $\rho_{app}(D_j, F_i) = 0.65\rho(H_{j,u}, H_{i,u}) + 0.35\rho(H_{j,l}, H_{i,l})$ 
      if  $\rho_{app}(D_j, F_i) \geq T_{app}$  do
         $F_i = D_j \subset D$  break;
      end if
    end for
    if  $F_i \notin D$  do
      Add  $F_i$  to  $D$  as a new feature dimension
    end if
  end for
  for Each feature dimension  $D_k$  in dimension set  $D$  do
    if Frames in  $D_k$  outnumber the set threshold  $T_F$  do
       $D_k$  is determined to be a qualified feature dimension
    end if
  end for
  Count the number of qualified dimensions in dimension set  $D$ 

```

---

### 5.3. SVM-Based Object Matching

#### 5.3.1. Support Vector Machine

Support Vector Machine (SVM) is a powerful classifier which shows many special advantages in solving classification of nonlinear high-dimensional pattern recognition problems with small samples<sup>[15]</sup>. This paper adopts SVM to map from nonlinear space to a higher dimensional space and construct the optimal separation hyper-plane. More specifically, SVM is used to fuse upper and lower appearance features in training data.

Given a set of labeled samples  $\{x_i, y_i\}, i = 1, 2, \dots, n$ ,  $x_i$  is an  $m$ -dimensional vector,  $y_i \in \{1, 0\}$  is the topic label. The samples labeled with  $y_i = 1$  belong to positive class while the samples labeled with  $y_i = 0$  belong to negative class. After training the known sample data, a classifying function is generated as formula

$$f(x) = \text{sgn}\{W^T X + b\}, \quad (16)$$

The output of  $f(x)$  determines the class of input vector  $X$ . After being mapped by nonlinear projection, the samples become linearly separable. According to [16], a proper kernel function  $K(x_i, x_j)$  can help to classify the mapped data. Then there is an optimal separating function

$$\max O(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j), \quad (17)$$

where,  $\alpha_i \geq 0, i = 1, 2, \dots, l$  is Lagrange multiplier. Consequently, the classifying function becomes

$$f(x) = \text{sgn}\left\{\sum_{i=1}^n \alpha_i^* y_i K(x_i, x) + b^*\right\}, \quad (18)$$

Here,  $\alpha_i^*, b^*$  represent the optimal Lagrange multiplier and the corresponding classifying threshold. This paper employs Gaussian radial function as a kernel function as following

$$K(x_i, x_j) = \exp\left\{-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right\}, \quad (19)$$

### 5.3.2. SVM Training and Testing

The SVM is trained manually by the source frames of tracked object collected before-hand at the beginning of tracking process. When the first camera comes, all its inner frames are tested by the trained SVMs to position the tracked object. The frames of the tracked object captured in the first camera will train new SVMs and replace the old SVMs to test the new coming frames in its next camera. The training stage will continue until the tracked object appears in principal view. The SVMs trained in principal view are the final SVMs and all the frames in later cameras will be tested by the final SVMs thereafter. Each of appearance feature dimensions in principal view(or in interim principal view) corresponds to a  $SVM_j$  ( $j = 1, 2, \dots, N_{hist}$ ).

For an object candidate set  $set\{O_i\}$ , the similarity of each object candidate  $O_i$  with the tracked object  $O_0$  is computed. If  $\rho_{biom}(O_i, O_0) \leq T_{biom}$ , then  $O_i$  will be canceled from  $set\{O_i\}$ . The remaining  $m_{obj}$  candidates render test frames to  $SVM_j$ . For each candidate, if at least one of  $SVM_j$  outputs 1, the candidate would be recognized as the tracked object and will be contained in a red bounding box with the tracked label. If all the outputs are zero, the candidate would not be the tracked object.

## 6. EXPERIMENTAL RESULTS AND DISCUSSION

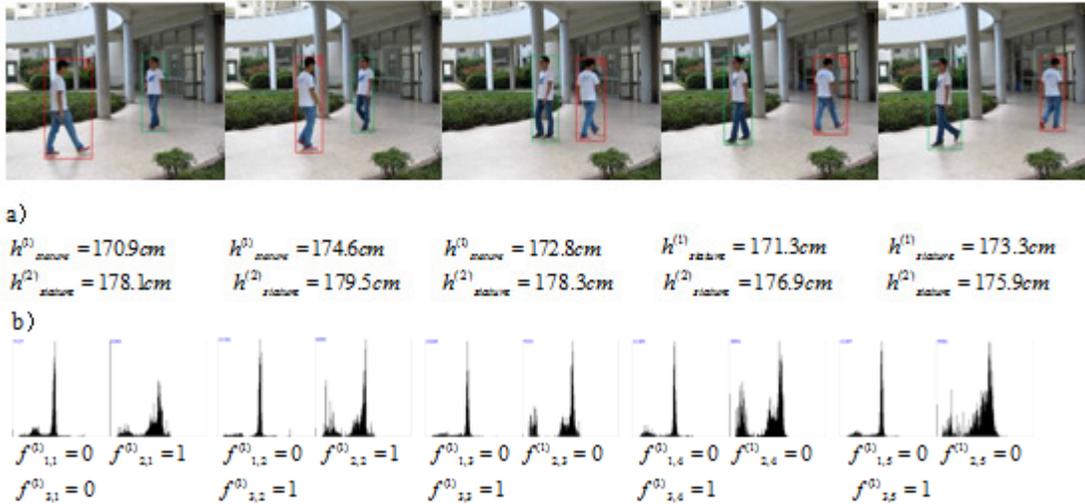
### 6.1. Experimental Setup

Numerous experiments are carried out to verify the feasibility and effectiveness of this proposal. The shooting facilities include two types of camera (Canon Power Shot A3200 IS, Canon PC1356) and they are installed in fixed height in 7 non-overlapping views. The testing platform is under VC++ and OpenCV 2.3.1 with Intel Core i7 870(2.93 GHz) CPU, 8 GB RAM and Windows 7 64bits OS.

The experiments are conducted in simple scenario and complex scenarios. In simple scenario, there exists only one or two moving objects and their moving trajectories are relatively regular. The simple scenario satisfies the single variable requirement of experiment and facilitates a detailed analysis of the advantages in this proposal. The complex scenario is composed of 7 non-overlapping views and there exists several individuals in field of views where may witness many mutual occlusions.. The experimental complex scenarios reflect closer to real-life situations and reveal the toughness of video surveillance. Before the tracking process starts, the SVM is trained in advance by the manually selected source frames of tracked person.

## 6.2 Experimental Results in Simple Scenario

An experiment in simple scenario where exists a short individual (object 1) and a tall individual (object 2) in same uniforms is initially conducted (see Fig.2). Each individual's geometrical parameters are computed statistically through the first 30-50 frames. Fig.2 a) lists the heights of Object 1 and Object 2 in frames within two rows respectively. Table 1 indicates that Object 2 will cease to be analyzed because its geometrical similarity with the tracked object is 0.68, less than threshold 0.8. Object 1 will continue being analyzed since its geometrical similarity with the tracked object is 0.92, more than 0.8. The sequences of Object 1 will be tested later by the SVMs which are generated through source sequences. Fig.2 b) exhibits the HSV histograms in frame sequences of Object 1 and the outputs of SVMs indicate the match result of each sequence.



**Fig.2** Object tracking of individuals with different geometrical features.  $h_{stature}^{(0)}$  in a) gives stature of object 1,  $h_{stature}^{(2)}$  gives stature of object 2. The left half of HSV histogram in b) stands for the upper appearance feature and the right half of HSV histogram stands for the lower appearance feature.  $f_{i,j}^{(k)}$  is the testing output of the  $k_{th}$  object from  $i_{th}$  SVM in frame  $j$ .

**Table 1** Geometrical parameters of moving objects

	Stature/cm	$\mu_{\Delta}$ /cm	$v_{\Delta}$ /cm	$F_{periodic}$ /frames	Similarity
Tracked object	174.2	0.320	0.021	13	---
Object 1	174.6	0.351	0.019	13	0.92 > 0.8
Object 2	179.5	0.392	0.020	15	0.68 < 0.8

Table 2 demonstrates the slump of erroneous matching rate with the increase of the geometrical components. Comparing 2nd row with 3rd and 4th row, the omit matching rate declines obviously while the matching rate has slight drops.

**Table 2** Matching parameters of different geometrical feature groups

Group	Matching rate(%)	Erroneous matching rate(%)	Omit matching rate(%)
None	95.8	8.2	8.2
$h_{stature}$	93	14.2	18.1
$(h_{stature}, (\mu_{\Delta}, v_{\Delta}))$	92.4	4.7	12
$(h_{stature}, (\mu_{\Delta}, v_{\Delta}), F_{periodic})$	91.2	0.6	9.1

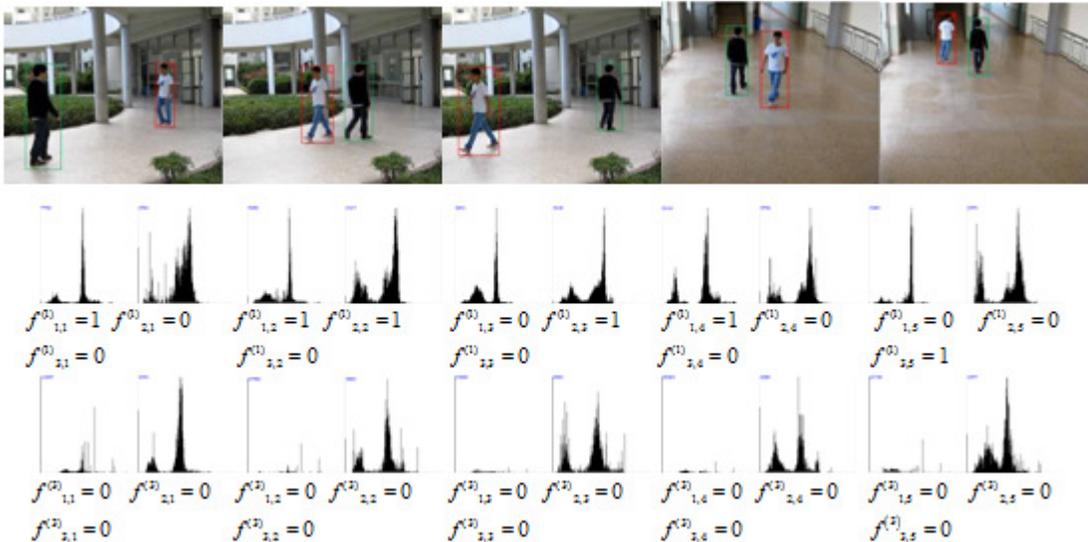
Note: Matching rate refers to the proportion of frames in which object matching occurs, it includes correct matching and incorrect matching; Erroneous matching rate refers to the proportion of frames in which the non-tracked object is falsely matched as the tracked object; Omit matching rate refers to the proportion of frames in which the tracked object is not matched with.

Another experiment in simple scenario is conducted where exists two geometrically similar individuals whose clothes are in huge contrast. Table 3 lists their geometrical parameters. The similarities of the two objects with the tracked object are both greater than the set threshold 0.8, and both objects would continue being analyzed. In the following process, appearance feature is analyzed to discriminate each individual and the tracking result is indicated by the outputs of SVMs, as shown in Fig.3.

**Table 3** Geometrical parameters of moving objects

	Stature/cm	$\mu_{\Delta}$ /cm	$v_{\Delta}$ /cm	$F_{periodic}$ /frames	Similarity
Tracked object	174.2	0.320	0.021	13	---
Object 1	174.3	0.341	0.020	13	0.94 > 0.8
Object 2	173.5	0.332	0.025	13	0.87 > 0.8

The footages in Fig.3 show the success of human tracking in this experimental situation even across two camera views where the shooting angles and illumination conditions change vastly.



**Fig.3** Human tracking of individuals with different appearance features. The HSV histograms of the individual in white are arranged in the second row, the HSV histograms of the individual in black are arranged in the third row.

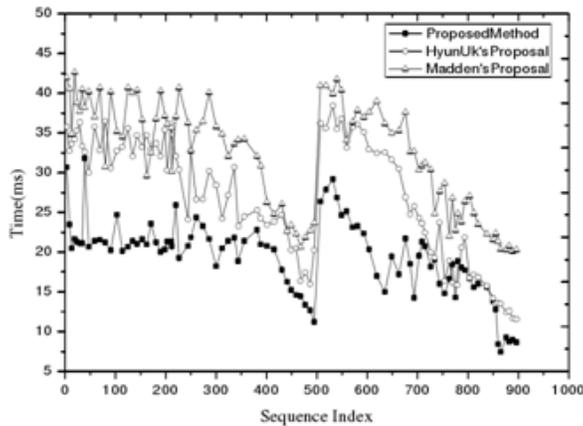
Table 4 lists the comparative matching parameters of this proposed method with Hyun-Uk’s proposal in [5]. The figures show the increase of accuracy rate and the decrease of omit matching rate.

**Table 4** Matching parameters of different human tracking proposals in simple scenario

Tracking proposal	Accuracy rate(%)	Erroneous matching rate(%)	Omit matching rate(%)
Proposed Method	93.3	0.6	2.8
Hyun-Uk’s in [5]	78.6	0.4	26.2

Note: Accuracy rate refers to the proportion of frames in which the tracked object is matched with accurately.

To verify the real-time of the appearance feature extraction in this paper, all the 898 frames captured from two views are preprocessed into 720×540 bmp images and their appearance features are extracted through the proposals of Hyun-Uk’s [5] and Madden’s [7] as well as this proposed method. The triple comparison diagram of computation time is illustrated in Fig.4.



**Fig.4** Comparison diagram of computation time of three different appearance feature extracting methods.

### 6.3 Experimental Results in Complex Scenarios

Another three experiments in complex scenarios are carried out to verify the feasibility of this proposal to track multi humans simultaneously. Seven cameras are installed in a wide area as shown in Fig.6 and none of them has pair-wise overlapping views. The view field of Camera 2 is manually selected as principal view.

When a moving object is first detected, its geometrical and appearance features will be stored in the system and it will be contained in a green bounding box. Fig.5 shows the tracks of object1, object2 and object5 in non-overlapping views.



Fig.5 The tracks of object 1,2,5 in five non-overlapping views exhibited in Row1, Row2 and Row3 respectively. The caption underneath indicates the image sequence index and camera belonging.

Note: The aspect ratio of the footages above may be reassigned for layout alignment.

The proposed method is further compared with other human tracking methods like Hyun-Uk's [5] Madden's [7] and Trevor's [11]. Table 5 lists the matching parameters of each method. The figures in first column not only reveal the decrease of accuracy rate when compare with the simple scenario but also show that this proposal significantly outperforms other three tracking methods in terms of accuracy rate and erroneous matching rate. However, this proposal has a higher omit matching rate than [5] and [7], but much lower than [11].

Table 5 Matching parameters of different human tracking proposals in complex scenarios

Proposal	Accuracy rate(%)	Erroneous matching rate(%)	Omit matching rate(%)
Proposed Method	84.3	5.6	10.8
Hyun-Uk's [5]	62.6	45.4	6.2
C. Madden's [7]	69	11	9.6
Trevor's [11]	77.6	8.1	23

According to temporal relation, it can retrieve the walking path of each tracked object. In this wide area, Object 1 walks along such a path: (North Gate, F1) to (North Hall, F1) to (Outdoor Corridor, F1) to (Left Stair, F2) to (Right Corridor, F2) as the blue route illustrated in Fig.6; Similarly, Object 2 walks along a path like the red route and Object 5 walks along a path like the yellow route.

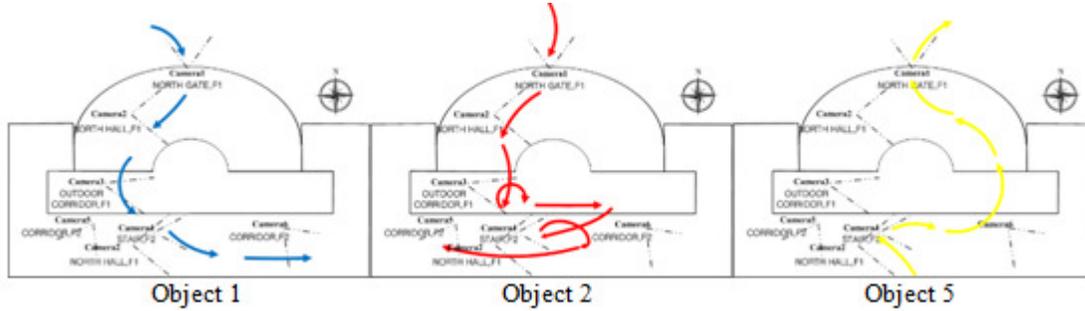


Fig.6 Walking path of each tracked object

#### 6.4. Discussion

Further analysis demonstrates the effectiveness of combining stature (height estimate),  $\mu_{\Delta}$ ,  $v_{\Delta}$  and  $F_{periodic}$  together in decreasing erroneous matching rate and omit matching rate. The obvious decline in term of omit matching rate from 2nd row to 3rd and 4th row in Table 2 means that the over filtering problem resulting from foreground errors could be suppressed by adding  $\mu_{\Delta}$  and  $v_{\Delta}$  into geometrical feature group. Over filtering problem is referred to the false filtering of capable object candidates in this paper.

The figures in Table 4 show the advantage of this proposed method in elevating the accuracy rate and reducing the omit matching rate. This attributes to the adoption of APVSA because it is designed to counter the appearance change in different visual angles. Whereas, since [5] extracts a global appearance feature rather than local major appearances, it outperforms this proposal slightly in term of erroneous matching rate. Fig.4 illustrates a triple comparison diagram of computation time. The contrast lines demonstrate the advantage of this proposal in computation time consumption. Comparing with [7], the proposed method benefits from the HSV uniformly - partitioning method as well as the choice of local appearance features. Furthermore, the proposed method is more applicable than [13] owing to the flexibility of the partitioning method.

The decrease of accuracy rate in complex scenario comparing with simple scenario is mainly due to the mutual occlusions of objects that occur frequently in complex scenario. But figures in Table 5 also show that this proposal significantly outperforms other tracking methods in terms of accuracy rate and erroneous matching rate. It attributes to the use of height estimate difference parameters  $\mu_{\Delta}$ ,  $v_{\Delta}$  and walking periodicity  $F_{periodic}$  in object filtering operation as well as the design of APVSA to maximize appearance feature dimensions of the training samples. The fusion of  $\mu_{\Delta}$ ,  $v_{\Delta}$  and  $F_{periodic}$  with height estimate strengthens the robustness to height disturbances. The adoption of APVSA strengthens the robustness to visual angle change. However, this proposal performs worse than [5] and [7] in term of omit matching rate, but much better than [11]. This is mainly due to the over filtering problem caused by foreground extraction errors.

The numeric results and further analyses in these experiments demonstrate the superior performance of this proposal in disjoint human tracking tasks. This proposal moves the tracking

methods forward as it can counter the illumination, visual angle change simultaneously as well as cut the time consumption of appearance feature extraction.

## 7. CONCLUSION

This paper presents a robust method to re-identify human object across non-overlapping views based on appearance and geometrical features. All the features extracted in this paper are intended to keep unchanged in different views. This proposal mainly benefits the disjoint human tracking in three aspects:

- 1, The local uniformly-partitioned HSV color features are extracted in a real-time speed which manages to resist the illumination change as well as cut computation time consumption in appearance feature extraction phase.
- 2, A new geometrical feature which integrates height estimate, height estimate difference parameters and the sinusoidal periodicity of walking is created with avoidance of the complexity of extraction and can robustly counter the visual angle change even across disjoint cameras.
- 3, An innovative matching method is presented based on the designs of APVSA, AFDDA and the adoption of SVM. APVSA facilitates the robustness of people re-identification to appearance change in different visual angles and SVM strengthens the discrimination of human individual.

Experimental results in simple and complex scenario indicate the accuracy and efficiency of this proposal. Further results demonstrate the effectiveness of the combined geometrical features in reducing the erroneous matching rate and the effectiveness of APVSA in decreasing omit matching rate. In spite of the abovementioned advantages, this proposal cannot overcome the occlusion problem which often occurs in complex scenario. The future work will focus on this.

## ACKNOWLEDGEMENTS

We thank the reviewers for helping us to improve this paper. This work is supported by Startup Project for Introduced Talents of Sichuan University.

## REFERENCES

- [1] Olivier Barnich, Marc Van Droogenbroeck, ViBe: A Universal Background Subtraction Algorithm for Video Sequences, *transactions on image processing*,.20(.6) (2011) 1709-1724.
- [2] Yao Lu , Ling Li , Patrick Peursum ,Human Pose Tracking Based on Both Generic and Specific Appearance Models, 12th International Conference on Control, Automation, Robotics & Vision,2012,p.1071-1076
- [3] T. D'Orazio, P.L.Mazzeo, P.Spagnolo, Color Brightness Transfer Function Evaluation for Non-overlapping Multi Camera Tracking, Third ACM/IEEE International Conference on Distributed Smart Cameras, 30 August-2 September,2009.
- [4] Gianfranco Doretto , Thomas Sebastian, Peter Tu, Jens Rittscher, Appearance-based person re-identification in camera networks: problem overview and current approaches, *Journal of Ambient Intelligence and Humanized Computing* 2(2) ( 2011) 127-151.

- [5] Hyun-Uk Chae, Kang-Hyun Jo, Appearance Feature Based Human Correspondence under Non-overlapping Views, *Emerging Intelligent Computing Technology and Applications*, 5754 ( 2009) 635-644.
- [6] O. Javed and K. Shafique and M. Shah, "Appearance Modeling for Tracking in Multiple Non-Overlapping Cameras," *IEEE Conference on Computer Vision and Pattern Recognition*, 2 (2005) 26-33.
- [7] C. Madden , M. Piccardi. A Framework for Track Matching Across Disjoint Cameras using Robust Shape and Appearance Features, *IEEE Conference on Advanced Video and Signal Based Surveillance*, 5-7 September 2007.
- [8] Yu-Chih Lin, Yu-Tzu Lin, Human recognition based on plantar pressure patterns during gait, *Journal of Mechanics in Medicine and Biology*,13(2),2013
- [9] Takayuki Hori ; Jun Ohya ; Jun Kurumisawa, Identifying a walking human by a tensor decomposition based approach and tracking the human across discontinuous fields of views of multiple cameras, *Computational Imaging VIII*, 75330 ( 2010).
- [10] YuChih Lin, BingShiang Yang, Yu Tzu Lin,Yi Ting Yang, Human Recognition Based on Kinematics and Kinetics of Gait, *Journal of Medical and Biological Engineering*, 31(4) (2010) 255-263.
- [11] Trevor Montcalm, Bubaker Boufama, Object Inter-camera Tracking with non -overlapping views: A new dynamic approach, *2010 Canadian Conference Computer and Robot Vision*,June,2010, 354-361.
- [12] Chenhui Yang, Weixiang Kuang, Robust Foreground Detection Based on Improved ViBe in Dynamic Background. *International Journal of Digital Content Technology and its Applications (JDCTA)* 7(4) (2013) 754-763.
- [13] Muhammad Riaz, Gwangwon Kang, Youngbae Kim, Sungbum Pan, and Jongan Park, Efficient Image Retrieval Using Adaptive Segmentation of HSV Color Space. *International Conference on Computational Sciences and Its Applications ICCSA 2008*, vol.55 (2008) 491 – 496.
- [14] Walking. Williams and Wilkins , *Human Walking*. (V.Inman, H.J.Ralston ,and F.Todd. ,1981).
- [15] Dawei Li, Lihong Xu, Erik D. Goodman, Yuan Xu and Yang Wu, Integrating a statistical background-foreground extraction algorithm and SVM classifier for pedestrian detection and tracking. *Integrated Computer-Aided Engineering*, 20 (2013) 201–216.
- [16] John Wiley & Sons, New York, *Statistical learning theory*. ( Vapnik, V.N.,1998).

## AUTHORS

**Binbin Liang** was born in 1990. He got a Master Degree in Engineering from the College of Civil Aviation at Nanjing University of Aeronautics and Astronautics. His research interests include civil aviation emergency management and computer vision.



**Dr. Songchen Han** was born in 1964. He obtained his Ph.D. in Engineering at Harbin Institute of Technology. He is currently a professor in Sichuan University in China. His research interests include (1) next generation air traffic management system (2) air traffic planning and simulation.



**Yan Zhu** was born in 1991. She is studying for a Master Degree in Engineering from the College of Civil Aviation at Nanjing University of Aeronautics and Astronautics. Her research interests include civil aviation emergency management and general aviation emergency rescue.



**Liping Di** was born in 1964. She obtained his Bachelor Degree in Aircraft Control at Harbin Institute of Technology, China. She is currently a teacher in Sichuan University in China. Her research interest is aircraft control.

