Neural Comput & Applic
DOI 10.1007/s00521-015-1900-5

ORIGINAL ARTICLE

Fuzzy qualitative human model for viewpoint identification

Chern Hong Lim[1] · **Chee Seng Chan**[1]

Received: 6 December 2014 / Accepted: 24 March 2015

Abstract The integration of advance human motion analysis techniques in low-cost video cameras has emerged for consumer applications, particularly in video surveillance systems. These smart and cheap devices provide the practical solutions for improving the public safety and homeland security with the capability of understanding the human behaviour automatically. In this sense, an intelligent video surveillance system should not be constrained on a person viewpoint, as in natural, a person is not restricted to perform an action from a fixed camera viewpoint. To achieve the objective, many state-of-the-art approaches require the information from multiple cameras in their processing. This is an impractical solution by considering its feasibility and computational complexity. First, it is very difficult to find an open space in real environment with perfect overlapping for multi-camera calibration. Secondly, the processing of information from multiple cameras is computational burden. With this, a surge of interest has sparked on single camera approach with notable work on the concept of view specific action recognition. However in their work, the viewpoints are assumed in a priori. In this paper, we extend it by proposing a viewpoint estimation framework where a novel human contour descriptor namely the fuzzy qualitative human contour is extracted from the fuzzy qualitative Poisson human model for viewpoint analysis. Clustering algorithms are used to learn and classify the viewpoints. In addition, our system is also integrated with the capability to classify front and rear views. Experimental results showed the reliability and effectiveness of our proposed viewpoint estimation framework by using the challenging IXMAS human action dataset.

Keywords Human motion analysis · Video surveillance system · Computer vision · Fuzzy qualitative reasoning

1 Introduction

Video camera nowadays has become inexpensive and affordable for the consumer applications. With the integration of advance technologies such as human motion analysis (HMA) into the video camera system, it provides the capability to understand the human behaviour automatically from the video [1, 5, 8, 17, 21]. Such system has raised the interest of the community to implement the intelligent video surveillance system [12, 13, 29]. To achieve this, a robust intelligent video surveillance system with no constraint on the person viewpoint is required. Consequently, view-invariant HMA system is getting more attention from the researchers [11], as in natural, human has no restriction to perform an action at a fixed camera viewpoint and this kept away the conventional view-dependent HMA systems [1]. Many works have been reported in view-invariant HMA, but most of the algorithms require the information obtain from multiple cameras during their processing such as the image features (e.g. shape) from every camera for 3D human model reconstruction. This is an impractical solution for two reasons. In a real-world environment, it is a daunting task to locate a place with good overlapping for multi-camera calibration. Furthermore, the processing using numbers of camera is very computational expensive.

✉ Chern Hong Lim
ch_lim@siswa.um.edu.my

Chee Seng Chan
cs.chan@um.edu.my

[1] Center of Image and Signal Processing, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

Published online: 24 April 2015

In order to solve this, [15] presented a framework that introduces the concept of performing multi-view action recognition within single camera and we denote it as view specific action recognition framework in our paper. The algorithm requires first identifying the viewpoint of a person captured from a video camera and followed by action classification using the corresponding viewpoint action model. However, in their work, the viewpoints are assumed in a priori. In this paper, we extend the work by introducing a viewpoint estimation framework as depicted in Fig. 1. In the process cycle, a human segment represented in silhouette image is extracted from the input image frame and normalized into the Fuzzy Quantity Space (FQS) [19] with aid of Poisson solution [9] to construct the fuzzy qualitative Poisson human model (FQ-PHM). Then the novel fuzzy qualitative human contour (FQHC) can be extracted from the FQ-PHM for viewpoint learning and estimation.

In summary, our contributions and its corresponding advantages are as follows: (1) We proposed the FQ-PHM that is invariant to human size and the body anatomy. In addition, the FQ-PHM standardizes the body landmarks on the FQS. This means ideally, the positions of the body parts are the same for every human in FQ-PHM. (2) A novel human contour descriptor called FQHC is generated from the proposed FQ-PHM, which is capable of distinguishing different human viewpoints. (3) An extra merit is given to our proposed framework by introducing a mechanism to estimate front and rear views of a person from a video camera, which is rather important but neglected by most of the researchers. As a summary, this framework can be integrated into the view specific action recognition framework [15] as a prerequisite.

The rest of the paper is organized as follows. Section 2 covers the related works in view-invariant HMA and viewpoint estimation. Our proposed FQ-PHM is discussed in Sect. 3 with the demonstration of the FQHC generation. Section 4 explains how we learn and estimate the viewpoints by using clustering algorithms. Apart from that, the mechanism to distinguish between front or rear views of a person is presented. Section 5 discusses the experiment settings and results. Finally, we conclude with suggestions of future work in Sect. 6.

2 Related works

View invariant is one of the current trend in the research of vision-based HMA with numbers of recent published surveys [1, 10, 11, 17, 27]. However, most of the works focus

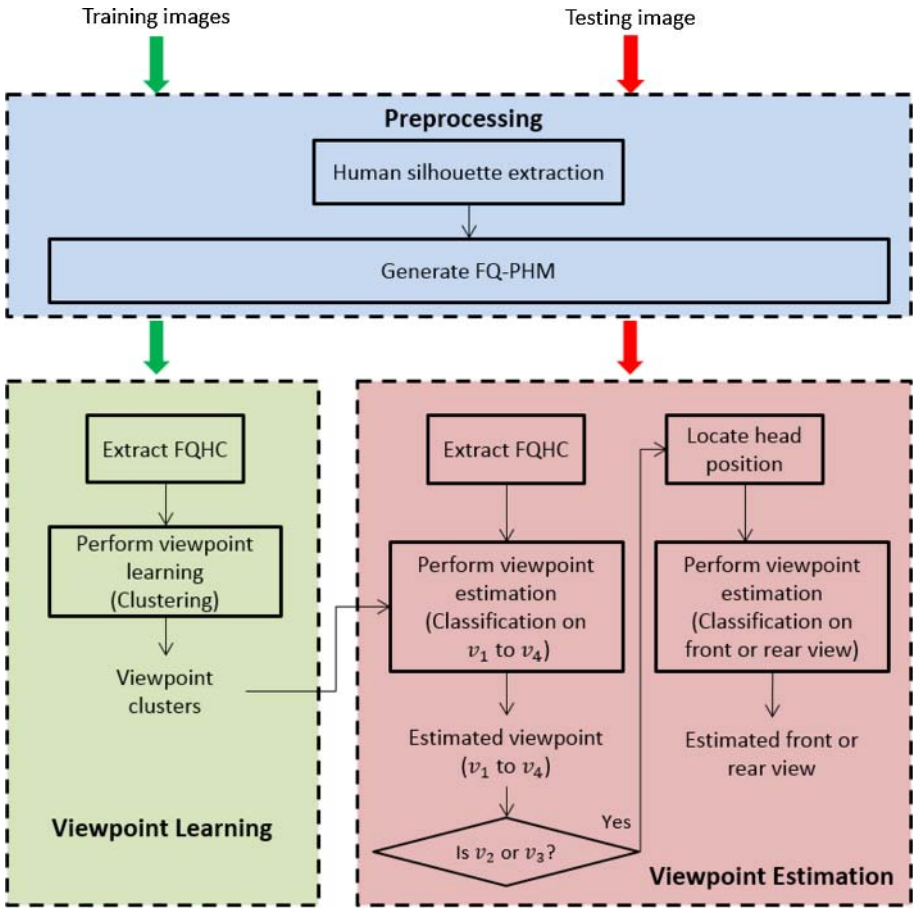

Fig. 1 The proposed viewpoint estimation framework

on the multiple cameras approaches where the information from different viewpoints is required to achieve view-invariant capability [2–4, 25, 26, 28]. In these approaches, they can be categorized into 3D human reconstruction methods and multi-view geometry approaches.

In 3D human reconstruction methods, the 2D human models obtained from multiple cameras are extended into 3D human models to reconstruct the human shape in the volumetric space. For instance, the voxel person [3] is introduced to reason elderly abnormal activity of falling with fuzzy logic approach. Weinland [26] proposes a new framework to model actions from multiple cameras by spatially integrated into visual hulls and accumulate into motion histogram volumes (MHV). View-invariant features are extracted from the Fourier space of MHV for action recognition.

On the other hand, epipolar geometry (the intrinsic projective geometry between two views) is popular nowadays to analyse body posture from cameras at different viewing angle for action recognition. For example, [4] utilizes the epipolar geometry to induces a fundamental matrix between two fixed cameras and the concept of fundamental ratios is investigated, which are invariant to camera intrinsic parameters in view-invariant action recognition. Besides that, [28] introduces the extension of standard epipolar geometry to the geometry of dynamic scenes where the cameras are moving to study the action.

Although the above-mentioned approaches achieved significant results in view-invariant action recognition, one of the drawbacks of using multi-camera approach is that it is only applicable to closed-controlled environment and it is impractical and expensive to deploy in a real-world environment. [15] remedies these limitations by proposing a framework to model human action from multiple viewpoints and perform recognition within a single camera. However, the viewpoints are assumed to be known in advance. In this single camera view-invariant HMA system, the person viewpoint information is very important as the pattern of similar action performed from different viewpoints varies. [23] verified this in their work and further showed that the better viewpoints are those where the action is easy to recognize. Thus, they propose three measurements (spatial, temporal and spatial-temporal) to detect a good viewpoint where the action is recognizable. In their experiment, they show that the selection of viewpoint does improve the action recognition rate. Inspired from this work, we understand the importance of analysing the different viewpoints correspond to the human actions. However, in the real world, a person is not always appearing at the best viewpoint in front of the camera. In our work, we take into consideration of all the viewpoints for different actions instead of just selecting the best viewpoints for corresponding human action as in [23].

In what constitutes the closer work to ours, [22] proposed a methodology to perform viewpoint estimation from the man-made environments to infer human activity. First, they model the body poses and silhouettes using a reduced set of training viewpoints. Second, they exploit projective geometry transformation for shape registration to improve silhouette-based pose estimation of the subjects from the real environment. Homography transformation is then employed to align between the testing viewpoint to the corresponding training viewpoint. Their work is currently tested only on the walking activity with subjects that have almost similar shape variation. Their methodology might suffer from the inconsistency in different human anatomy such as size, height, and pose. What differentiates our work from theirs is we utilized the Poisson solution incorporate with FQS to yield a normalized fuzzy qualitative human model, which we name it as the FQ-PHM to handle the above-mentioned problems. Moreover, our proposed approach is implemented with the capability to learn and classify the viewpoints and also distinguish between front and rear views of a person.

3 Fuzzy qualitative Poisson human modelling

A sophisticated human modelling approach is very important to most of the HMA systems where it provides much valuable information such as features for the inference process. In this paper, we proposed to model the human body in a fuzzy qualitative manner called FQ-PHM. To begin with, we extract the silhouette image of a person from the input image. Secondly, we map the human silhouette into the FQS. However, a problem arises that, without proper normalization technique, the human model is irregular for every human subject in terms of scale and position due to the variations of human anatomy, e.g. body size, height and pose. Such variations will affect the performance of the human model in extracting the features. Following sections discuss on how we deal with this problems by incorporating Poisson solution into the construction of FQ-PHM and the extraction of the novel FQHC descriptor for viewpoint estimation.

3.1 Construction of FQ-PHM

The overall pipeline to construct FQ-PHM is shown in Fig. 2. As mentioned earlier, we provide a solution to achieve better normalization of human body by applying Poisson solution [9] and map it to the FQS. The purpose of using Poisson solution is to obtain the reference point, $r = \{x_{\text{ref}}, y_{\text{ref}}\}$ of the person, which serves as the origin to map to the FQS. In this paper, we define the reference point as the lower part of the torso of a human body, which is estimated through (1),

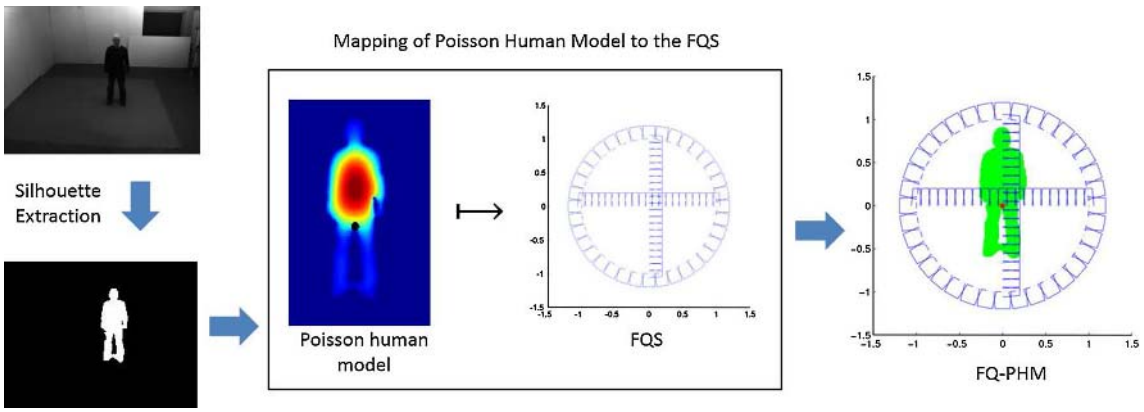

Fig. 2 The overall pipeline to generate FQ-PHM. This figure is best viewed with colour

$$
\begin{aligned}
[x_{\text{mid}}, y_{\text{mid}}] &= \max[U(x,y)] \\
x_{\text{ref}} &= x_{\text{mid}} \\
y_{\text{ref}} &= y_{\text{mid}} + (3 * (L/5))
\end{aligned} \tag{1}
$$

where $U(x,y)$ is the Poisson solution [9] of the human silhouette and L is the vertical length of the subject in the image. According to [15], this reference point can correctly designate the position of the body parts, especially the limbs. The principle behind is that value of U increases quadratically as it approaches to the centre. The level sets of U represent smoother versions of the bounding contour with the external protrusions, (where in human context, it refer to the limbs and head) disappearing at relatively low values of U. This is different from the distance transform, which smoothens the shape near concavities while introducing discontinuities near convex sections of the contour. Also unlike the distance transform in which every value is determined by a single contour point (the nearest), the values assigned by the Poisson equation take into account many points on the boundaries and so they reflect more global properties of the silhouette. In human representation, this is giving prudent information as ideally the highest value from the Poisson solution is at the middle of the torso part (highest intensity of red colour in Poisson human model in Fig. 2). This is because in general the torso is the largest part of human body. Here, it helps to retrieve the lower torso part with (1) and serves as a reference point to map the human body with the normalized range of [0 1], B_{norm} into the FQS.

$$B_{\text{norm}} \mapsto \text{FQS} \tag{2}$$

The construction of FQ-PHM adopted the FQS that was introduced in [19]. It consists of translation components (x & y-axis) that are denoted as FQS_x and FQS_y and orientation component (unit circle) that is denoted as FQS_o with their implicit qualitative states, QS as represented in (3).

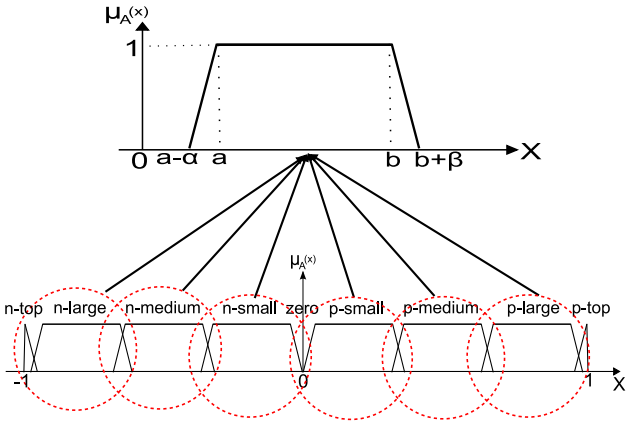

Fig. 3 4-Tuple qualitative state

$$
\begin{aligned}
\text{FQS}_x &= \{\text{QS}_x^1, \text{QS}_x^2, \ldots, \text{QS}_x^M,\} \\
\text{FQS}_y &= \{\text{QS}_y^1, \text{QS}_y^2, \ldots, \text{QS}_y^M,\} \\
\text{FQS}_o &= \{\text{QS}_o^1, \text{QS}_o^2, \ldots, \text{QS}_o^N,\}
\end{aligned} \tag{3}
$$

These components of FQS utilized a parametric approximation of the membership function to represent the qualitative state where the membership distribution of a normal convex fuzzy number is approximated by the 4 tuples, $\text{QS} = [a, b, \alpha, \beta]$ [24]. In the recent trends, 4-tuple fuzzy numbers have been utilized in constructing the qualitative states [6, 16, 19] that endowed with the capability to model the uncertainties.

The qualitative state is constructed from the 4 tuples as shown in Fig. 3, in this way, when there does exist a precise qualitatively distinct landmark value, this value can also be represented in the form of a 4-tuple fuzzy number. Furthermore, even if the landmarks are only partially known, say, in terms of the lower and upper boundaries of the range within which a landmark value falls, such

knowledge can still be encoded by the 4-tuple version of a real interval.

In this work, we empirically choose the total number of qualitative state in the translation components as $M = 21$ and total number of state in the orientation components to be $N = 36$ but not limited to. The number of qualitative states is application dependent. In general, too many the qualitative states will increase the computational complexity and also causes the overfitting problem. On the other hand, too less the qualitative states will reduce the effectiveness of the overall system as the fuzzy interval is too broad and might be unable to obtain the vital information of the human parts. To understand it further, [14] provides better explanation on the effect of different qualitative states. In fact, such representation outperforms the crisp Cartesian and unit circle by its capability of relaxing the uncertainties abounded in the processing [6, 14, 16, 18, 24].

In addition, it is worth to highlight that the advantages of the FQ-PHM is scale invariant and position standardized (Fig. 4). The normalization step of the human body utilizing the reference point allows the fixation of the human body parts to the FQS. For example, the head is always fall into $\mathrm{FQS}_x = \{\mathrm{FQS}_x^{10}, \mathrm{FQS}_x^{11}, \mathrm{FQS}_x^{12}\}$ and $\mathrm{FQS}_y = \{\mathrm{FQS}_y^{2}, \mathrm{FQS}_y^{3}, \mathrm{FQS}_y^{4}\}$ as can be observed from Fig. 4. From the figure, we can conclude that the proposed FQ-PHM is able to normalize the different human anatomy such as body sizes, heights and poses. In the next step, we will demonstrate the extraction of FQHC descriptor from the proposed FQ-PHM for viewpoint learning.

3.2 Fuzzy qualitative human contour descriptor

With years of study about human vision in cognitive science, neuropsychology, and neurophysiology, the researchers agreed with the argument that human recognizes an object from its appearance such as the contour, texture, and colour information [20]. Many researches in computer vision are inspired by this with the notable work in human detection, and histogram of gradient (HOG) descriptor is introduced in [7]. To the extend, in this paper, we extract human contour descriptor from the proposed FQ-PHM using the distance computation from the reference point towards human edges as demonstrate in Fig. 5.

We first obtain the set of outer edge pixels, e_j, of the human model for every QS_o and denoted as $P_{\mathrm{QS}_o^n}$,

$$P_{\mathrm{QS}_o^n} = \{e_1, e_2, \ldots, e_J\} \in \mathrm{QS}_o^n \tag{4}$$

where $n = 1, 2, \ldots, N$. We then compute the average of the distances from the reference point towards all the edge pixels that are bounded in the corresponding QS_o and designated as $\bar{D}_{\mathrm{QS}_o^n}$,

$$\bar{D}_{\mathrm{QS}_o^n} = \frac{1}{\left|P_{\mathrm{QS}_o^n}\right|} \sum_{j=1}^{J} \left\|e_j - r\right\|^2 \tag{5}$$

Next, a descriptor for the human contour is constructed and represented as feature vector, s (6) with clockwise direction as shown in the right image of Fig. 5a. The dimension of s is determined by the resolution of the orientation component, N in the FQS, where here we use $N = 36$.

$$s = \left\{\bar{D}_{\mathrm{QS}_o^1}, \bar{D}_{\mathrm{QS}_o^2}, \ldots, \bar{D}_{\mathrm{QS}_o^N}\right\} \tag{6}$$

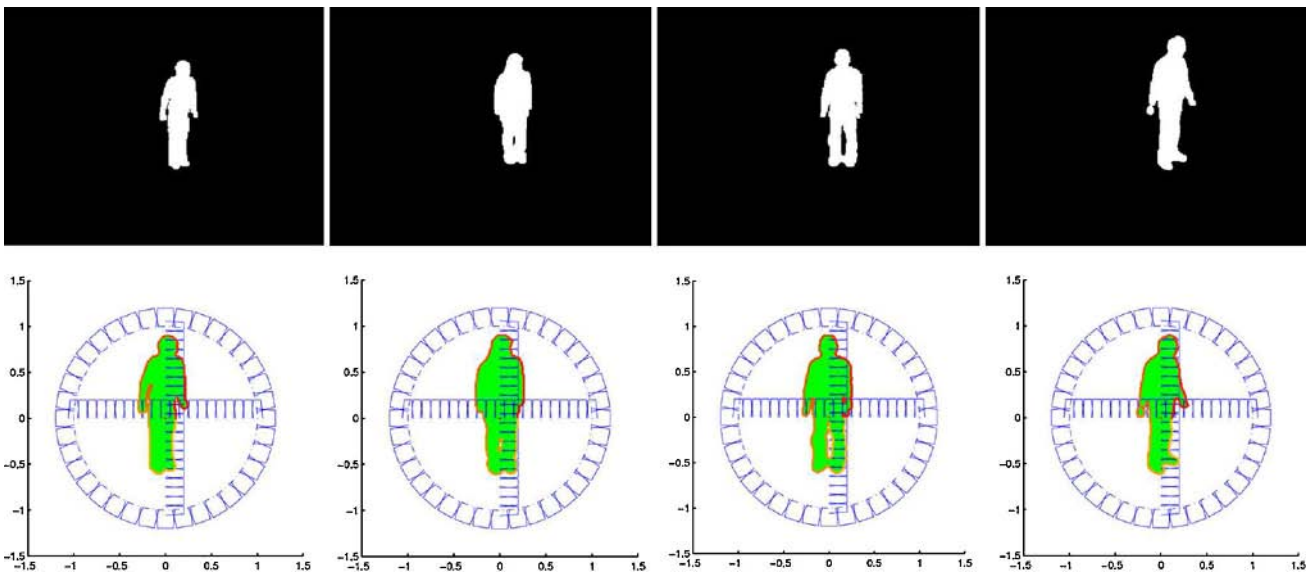

Fig. 4 *First row from left to right* represents the human silhouette images from the size of small to large. Second row shows the corresponding FQ-PHM. From the figure, one can notice that the size and the position of the body parts are almost similar for all the human subjects once they are being normalized onto the FQS with the reference point as the origin

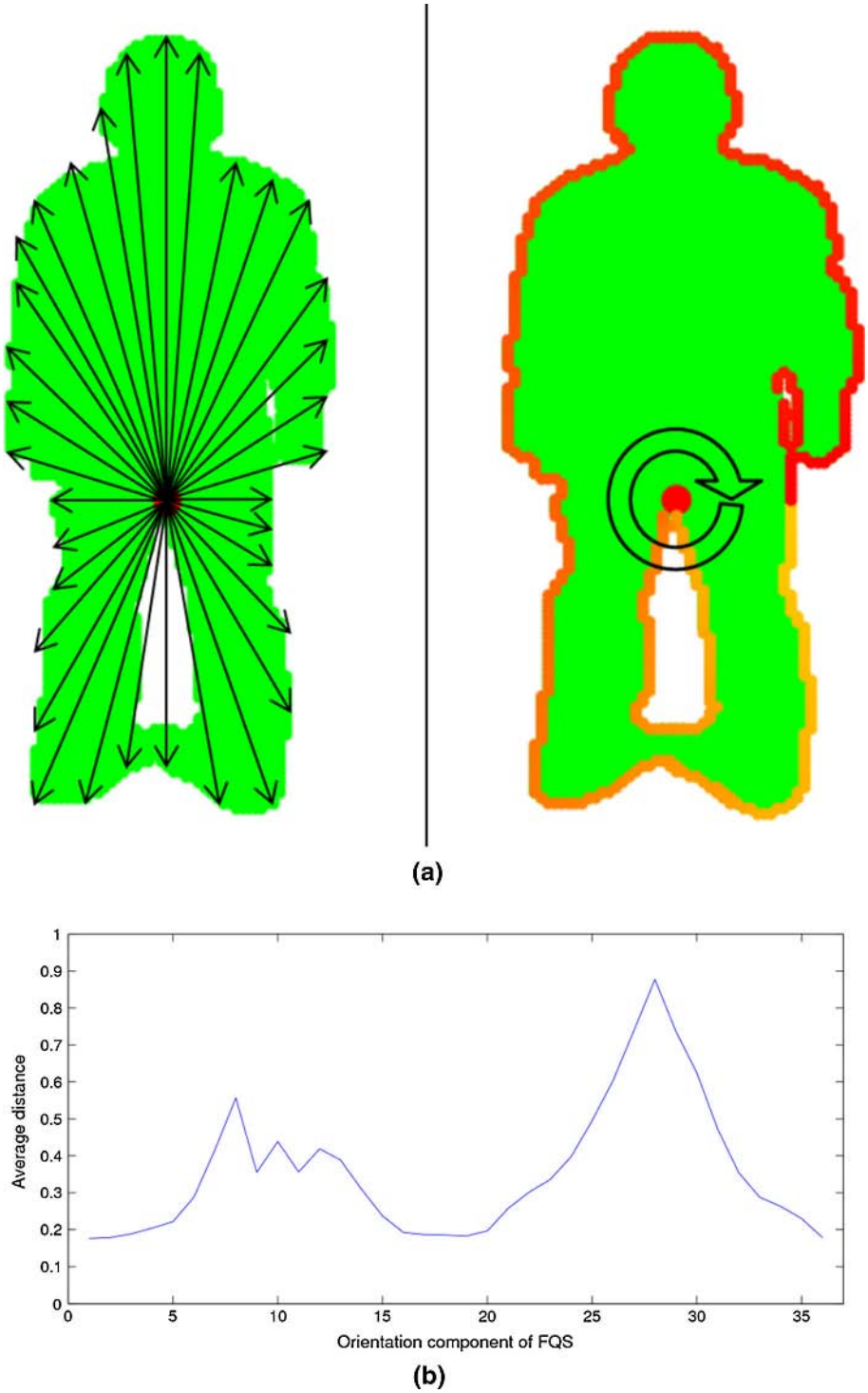

Fig. 5 **a** In the *left image*, we compute the distance from the reference point to the outer edge. The distance is organized according to clockwise direction as shown in the *right image*. **b** The example of the output of FQHC descriptor by averaging the distance in each orientation states of the FQS

One of the reasons of using averaging distance with bounded FQS_o is to create the descriptor with feature vector of fix dimension. By comparing it with directly using the edges information from the human silhouette, the feature vector can be in various dimensionalities due to the different sizes of every human subject. This causes the difficulty for the viewpoint learning in the next step. Besides that, the averaging of the edge pixels in the qualitative states can smoothen the inconsistency of some edge pixels due to the noise for better interpretation. The extracted human contour descriptors will then use to learn the viewpoints and generate the viewpoint classifier, which will be discussed in the next section. For better understanding, the the construction of FQ-PHM and the extraction of FQHC is summarized in Algorithm 1.

Algorithm 1 FQHC EXTRACTION FROM FQ-PHM

Require: An input image
Step 1: Silhouette extraction. Perform silhouette extraction to obtain binary representation of human body.
Step 2: Apply Poisson solution. Apply (1) towards the human silhouette image to obtain the reference point, r.
Step 3: Mapping to FQS. Normalize the human silhouette into the FQS with the range of [0 1] and r as the origin to obtain FQ-PHM.
Step 4: Extract human contour descriptor.
for all QS_o^n such that $1 \leq n \leq N$ **do**
  for all e_j such that $1 \leq j \leq J$ **do**
    Compute average of the distances, $\bar{D}_{QS_o^n}$ from r to e_j using (5).
  end for
end for
return FQHC Descriptor, s

4 Viewpoint learning and estimation

In order to justify the capability of the extracted human contour descriptors in distinguishing different human viewpoints, we employed unsupervised learning method (clustering algorithm) in learning a set of predefined viewpoints. Besides that, the learning outcome can be used as the classifier for viewpoint estimation task. In addition, we proposed a mechanism to distinguish between front and rear views besides the predefined viewpoints of a person from a camera.

4.1 Viewpoint learning from human contour descriptors

From the visual inspection in Fig. 6, one can notices that the FQHC descriptors are different from one viewpoint to another. Preliminary from this observation, we presume that the FQHC descriptor extracted from the FQ-PHM possesses the capability to differentiate different viewpoints. In order to verify this, we employed the clustering algorithm such as K-means (KM) and Fuzzy c-means (FCM) to learn the viewpoints.

Depend on the objective of a system, reader can set their own number of cluster. In this work, the cluster means the different viewpoints. To make it simple, we adopted the set of viewpoints defined in [15] which are horizontal view (v_1), diagonal view (v_2), vertical view (v_3) and top view (v_4) as visualized in Fig. 7 in our framework. These are the most common viewpoints that will be encountered from a video camera. According to [22], these viewpoints are sufficient for viewpoints analyses. Therefore, the number of cluster is $K = 4$ corresponding to each types of viewpoints, $V = \{v_1, v_2, v_3, v_4\}$. With the collection of the human contour descriptors, $S = \{s_1, s_2, \ldots, s_T\}$ is extracted from the training images, clustering algorithms to learn the viewpoints clusters are used, and the outcomes are expected to be similar to Fig. 7 with each of the descriptor; s is correctly assigned to their corresponding viewpoint cluster.

The main objective of this work is to perform viewpoint estimation, and in order to achieve that, we need a classifier to infer the viewpoint from an input image. As aforementioned in this section, the outcome of the viewpoint learning can be used for the classification purpose, but not directly. This is because of the random initialization of the clustering algorithm that prevents us from knowing the corresponding cluster centre, $C = \{c_v, v \in V\}$ to each of the viewpoints, v. In order to make the outcome a classifier, we need to assign the correct viewpoint label to the cluster centres. To achieve this, we count the occurrence of the viewpoints in each cluster outcome by matching it to the ground truth. The maximum viewpoint is assigned as the corresponding label for that cluster centre. As a result, one can simply use C as the classifier to estimate the person viewpoint of a testing image with distance metric calculation such as $L2$-norm.

4.2 Front- and rear-view estimation

Although the viewpoint classifier learned from the previous section can estimate the human subject into its corresponding predefined viewpoints (v_1 to v_4), it is unable to

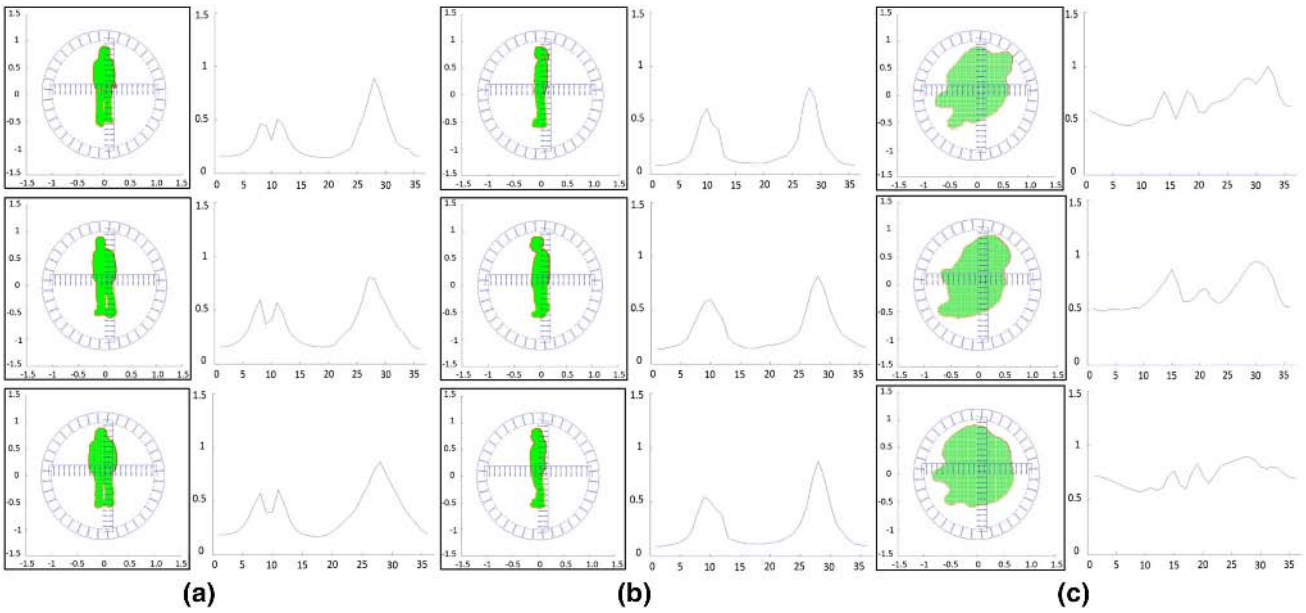

Fig. 6 Examples of different FQHC descriptor for **a** diagonal and vertical views, **b** horizontal view and **c** top view

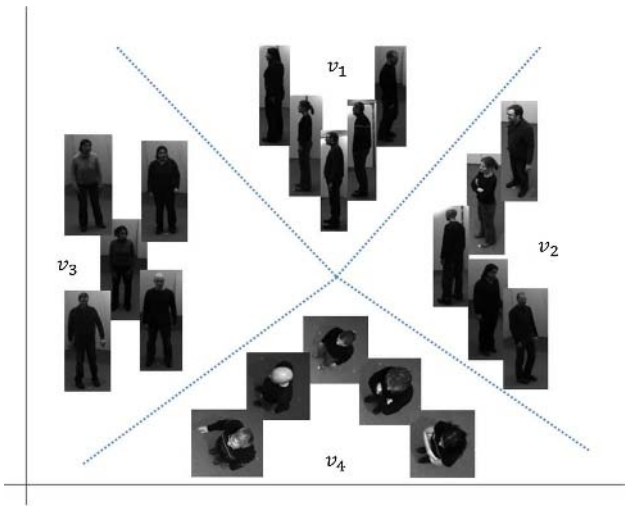

Fig. 7 Expecting learning outcome of the viewpoint clustering

determine whether a person is facing frontal or backward towards the camera. This is an additional cue for a HMA system in higher level interpretation such as behaviour understanding. For instance, the system can infer that action is hidden to a person who is facing backward to the camera and decided to increase the attention towards him or her and deduce the possible actions. Or on the other hand, the system might just ignore it to relax the computational cost. In this paper, we offer a solution to deduce front and rear views of a person by analysing the grayscale distribution of the human head as demonstrate in Fig. 8.

As a matter of fact, front- and rear-view estimation is only applicable to diagonal and vertical views (v_2 and v_3). Thus, in the viewpoint estimation process cycle, we first determine whether the person is facing diagonal, v_2, or vertical, v_3, views at the first stage. Then, we utilize the advantage of position standardized FQ-PHM to locate the head position of the person. This is done with human assistance but is only one time process. In this paper, based on the fixed number of the FQS translation components ($M = 21$), we estimate that the head of a person is mostly appear at $\mathrm{FQS}_x = \{\mathrm{FQS}_x^{10}, \mathrm{FQS}_x^{11}, \mathrm{FQS}_x^{12}\}$ and $\mathrm{FQS}_y = \{\mathrm{FQS}_y^{2}, \mathrm{FQS}_y^{3}, \mathrm{FQS}_y^{4}\}$. After we retrieved the head region, we obtain its grayscale histogram and compute the degree of homogeneity, H of the head using

$$H = 1 - \frac{\gamma(\vee q - \wedge q)}{\vee h - \wedge h} \tag{7}$$

$$q = \max(\mathrm{h})/5 \tag{8}$$

where $\vee h$ and $\wedge h$ are the lower boundary and upper boundary of the grayscale distribution of the head, similar to $\vee q$ and $\wedge q$ as the lower and upper boundary of the grayscale region after the alpha cut, q with (8), respectively. γ is the weighing parameter where we set $\gamma = 1.8$ to ensure that more weightage is put into the interpretation of the grayscale region ($\wedge q - \vee q$) as this is the dominant region that distinguishes between front and rear views in our case. The setting for alpha cut and the γ value can vary from one camera setup to another depend on the quality of image. Empirical solution is chosen for the best performance from the testing.

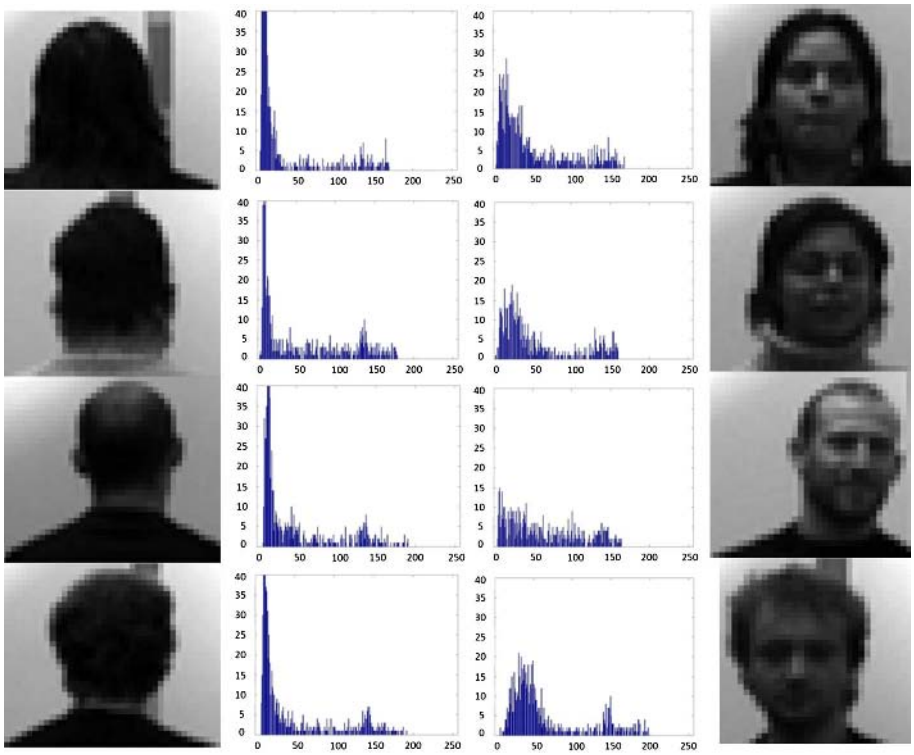

Fig. 8 Different grayscale distribution of front and rear views of human heads. The *first and second columns* show the rear view of human heads with theirs corresponding grayscale distributions, while *third and fourth columns* show the grayscale distributions corresponding to the front-view human heads. One can notice that the grayscale distribution of front views is more disperse than that of the rear views

Algorithm 2 VIEWPOINT LEARNING AND ESTIMATION

Require: Collection of human contour descriptors, S and a set of predefined viewpoints to determine the cluster number, K.

Step 1: Perform Clustering. Perform viewpoint learning by applying clustering algorithm to assign S to corresponding viewpoint cluster, C.

Step 2: Labeling clustering output. The arbitrary cluster centres from the learning outcome, C are labeled with the corresponding viewpoint by obtaining the maximum viewpoint label of each cluster with ground truth matching.

Step 3: View estimation. Perform viewpoint estimation by using C from the previous step as classifier.

if classified as v_2 or v_3 **do**

Locate the head position by utilizing FQ-PHM and compute the degree of homogeneity of the head using (7). Determine front or rear view with threshold, t.

end for

return Estimated viewpoint

The intuition of employing the degree of homogeneity here is, if a person is facing frontal to the camera, the degree of homogeneity of the person head is low because of the large grayscale dispersion caused by the inconsistent illumination from the face components (eyes, nose, ears and mouth). On the other hand, when a person is facing backward to the camera, the degree of homogeneity will become higher due to the consistent illumination from the hair or smooth skin which is assumed to have constant colour characteristic. Finally, front or rear view is deduced with a threshold, t. Algorithm 2 summarizes the steps from viewpoint learning to viewpoint estimation.

5 Experiment

In order to evaluate the reliability and the effectiveness of our proposed framework, we conducted two experiments where first is to test the discriminative strength of the FQHC descriptor in differentiate each predefined viewpoints and secondto test the effectiveness of the learned classifier and the front- and rear-view estimation algorithm. Both experiments employed IXMAS dataset that is publicly available at http://4drepository.inrialpes.fr/public/viewgroup/6. The predefined set of viewpoints from [15] (v_1, v_2, v_3 and v_4) is employed in the experiments with human annotated ground truth.

5.1 Discriminative strength of fuzzy qualitative human contour descriptor

The main objective of this experiment is to evaluate the capability of the FQHC descriptors that are extracted from the FQ-PHM in distinguishing one viewpoint from another. From the IXMAS dataset, we assume that the human subjects are in his or her initial standing position by extracting only the first five frames from each video captured by the cameras from all the viewpoints. We compared the performance of our proposed FQHC descriptor with HOG descriptor [7] by using KM and FCM as the clustering techniques, and the results are being evaluated in term of precision and recall. We perform 20 trials for each of the testing and the average results are reported in Table 1.

From Table 1, the precision and recall for v_1 and v_4 achieved high precision and recall but v_2 and v_3 are giving fair results. This is due to the confusion between the diagonal and vertical views as demonstrated in Fig. 9a. The human contour descriptors for v_2 and v_3 are found to be very similar to each other and thus induced confusions in the classification result. Although this is acceptable in the real environment, human has the difficulty to distinguish them too.

Table 1 Precision (Ps) and Recall (Rc) for the clustering results

	v_1		v_2		v_3		v_4	
	Ps	Rc	Ps	Rc	Ps	Rc	Ps	Rc
FQ-PHM_KM	**0.92**	0.74	0.57	0.57	0.58	0.72	**1.00**	0.99
FQ-PHM_FCM	0.83	**0.87**	**0.65**	**0.58**	**0.66**	**0.72**	**1.00**	0.99
HOG_KM	0.67	0.82	0.54	0.36	0.64	0.58	0.91	**1.00**
HOG_FCM	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail

Bold values indicates the best clustering result with the highest precisions and recalls

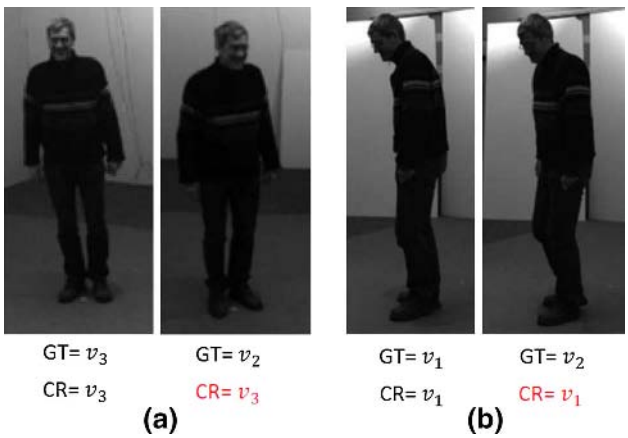

Fig. 9 The examples of confused viewpoints. Its ground truth is denoted as GT, and the computational result is denoted as CR. We can notice that the right image in **a**, the CR is conflict with GT where the computer miss classified it as v_3. While in **b**, the right image is miss classified as v_1. Although they are misclassification cases, but in human visual perspective, this is forgivable as both are very similar to their corresponding left images

Table 2 Error rate of the clustering results from 20 trials

	Error rate
FQ-PHM_KM	**0.2**
FQ-PHM_FCM	0.3
HOG_KM	0.8
HOG_FCM	1.0

Bold value indicates the best clustering result with the least error rate

Table 3 Precision and recall for the side and non-side classification

View	Precision	Recall
Side	0.61	0.67
Non-side	0.81	0.77

Due to the randomness of initialization in clustering algorithm, we are likely to obtain undesirable clustering outcome (which we defined it as fail case). However, this can be a good evaluation criterion to determine the discriminative strength of a descriptor. In common practice, the higher the success rate of a clustering algorithm to cluster the input data into the desire cluster, the better the discriminative strength of the descriptor is. Table 2 shows the error rates of the clustering results.

FQHC descriptor extracted from the FQ-PHM performed well with KM and FCM with low error rate but HOG receive high error rate in KM and even failed to perform clustering with FCM as shown in Table 1. From these results, we can conclude that not only selecting clustering technique is an important concern on the learning of the viewpoints but also higher attention should be put on the discrimination strength of the descriptor. From this experiment, FQHC descriptors that are extracted from FQ-PHM have achieved the objective with better reliability.

5.2 Evaluation on the viewpoint estimation

In this experiment, we tested two things. First, we use the cluster centres that learned from the previous experiment as a viewpoint classifier to estimate between side (v_1) and non-side (v_2, v_3) viewpoints using minimum $L2$-norm calculation. Secondly, we differentiate between front and rear views of the person from all the non-side views with our proposed mechanism. This is because side view does not valid for front- and rear-view estimation. In order to test the effectiveness and the robustness of our proposed method, we used different set of testing data, which consists of a total of 2250 image frames from the "turnaround"

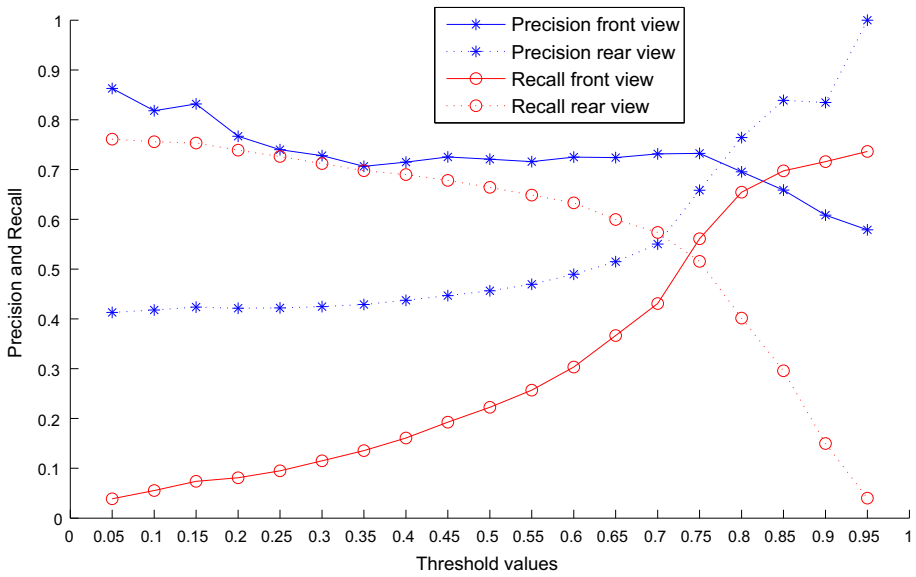

Fig. 10 Precision and recall of front- and rear-view estimation with different threshold values, t with $t = 0.75$ is achieving the best performance

action in IXMAS dataset (noted that the top view, v_4 is not included in this testing). The precision and recall for the side and non-side viewpoints estimation is shown in Table 3, while the result for front- and rear-view estimation corresponds to different threshold values, and t is presented in Fig. 10.

From Table 3, we obtained moderate results for side view but high precision and recall for the non-side view. Similarly, during the "turnaround" activity, there are frames that confused between side and non-side views because of slight movement of the person as shown in Fig. 9. Such confusions deteriorate the classification result.

As for the front- and rear-view estimation, Fig. 10 shows that the threshold value $t = 0.75$ is found to be the best threshold in the testing. However, this is varies from one setup to another, so reinitialize of the threshold value is necessary for a new camera setup. Conclusively, the estimation of viewpoints using the proposed algorithm and the estimation of front and rear views is achieving reasonable results.

6 Concluding remarks

We presented a novel human modelling methodology by incorporating the Poisson solution and the FQS yielding a new human model called the FQ-PHM that is proved to be invariant to different human anatomy (e.g. size, height, and pose) and is position standardized. In order to perform view estimation, FQHC descriptor is introduced that is extracted from FQ-PHM to represent the human contour, and its discriminative strength to distinguish different viewpoints has been tested with KM and FCM clustering algorithms. The experiments outcomes have shown their potential to be a classifier to infer different viewpoints. Moreover, we introduce a simple yet effective mechanism to determine front and rear views of a person by utilizing the FQ-PHM to locate the person's head with the measurement of its degree of homogeneity.

In this work, we achieve reasonable results for the experiments and apparently FQHC descriptor outperform the state-of-the-art HOG descriptor in viewpoint learning. However, the crisp classification method in the viewpoint estimation yields the confusion between the similar viewpoints, which will deteriorate the overall performance of the system. In future, we foresee that using fuzzy set theory to alleviate the crisp classification result can be a feasible solution to relax the above-mentioned problem.

Acknowledgment This research is supported by the Fundamental Research Grant Scheme (FRGS) MoE Grant FP027-2013A, H-00000-60010-E13110 from the Ministry of Education Malaysia.

References

1. Aggarwal J, Ryoo M (2011) Human activity analysis: a review. ACM Comput Surv 43(3):16:1–16:43
2. Ahmad M, Lee SW (2006) HMM-based human action recognition using multiview image sequences. In: International Conference on Pattern Recognition (ICPR), pp 263–266
3. Anderson D, Luke R, Keller J, Skubic M, Rantz M, Aud M (2009) Modeling human activity from voxel person using fuzzy logic. IEEE Trans Fuzzy Syst 17(1):39–49
4. Ashraf N, Shen Y, Cao X, Foroosh H (2013) View invariant action recognition using weighted fundamental ratios. Comput Vis Image Underst 117(6):587–602
5. Badi HS, Hussein S (2014) Hand posture and gesture recognition technology. Neural Comput Appl 25(3–4):871–878
6. Chan CS, Liu H (2009) Fuzzy qualitative human motion analysis. IEEE Trans Fuzzy Syst 17(4):851–862

7. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition (CVPR), vol 1, pp 886–893
8. El-Baz A, Tolba A (2013) An efficient algorithm for 3d hand gesture recognition using combined neural classifiers. Neural Comput Appl 22(7–8):1477–1484
9. Gorelick L, Galun M, Sharon E, Basri R, Brandt A (2006) Shape representation and classification using the poisson equation. IEEE Trans Pattern Anal Mach Intell 28(12):1991–2005
10. Holte MB, Tran C, Trivedi MM, Moeslund TB (2011) Human action recognition using multiple views: a comparative perspective on recent developments. In: Proceedings of the 2011 joint ACM workshop on human gesture and behavior understanding, pp 47–52
11. Ji X, Liu H (2010) Advances in view-invariant human motion analysis: a review. IEEE Trans Syst Man Cybern Part C Appl Rev 40(1):13–24
12. Lim MK, Tang S, Chan CS (2014) iSurveillance: intelligent framework for multiple events detection in surveillance videos. Expert Syst Appl 41(10):4704–4715
13. Lao W, Han J, de With PH (2009) Automatic video-based human motion analyzer for consumer surveillance system. IEEE Trans Consum Electron 55(2):591–598
14. Lim CH, Chan CS (2012) A fuzzy qualitative approach for scene classification. In: International conference on fuzzy systems (FUZZ), pp 1–8
15. Lim CH, Chan CS (2013) Fuzzy action recognition for multiple views within single camera. In: IEEE international conference on fuzzy systems (FUZZ), pp 1–8
16. Lim CH, Risnumawan A, Chan CS (2014) A scene image is nonmutually exclusive-a fuzzy qualitative scene understanding. IEEE Trans Fuzzy Syst 22(6):1541–1556
17. Lim CH, Vats E, Chan CS (2015) Fuzzy human motion analysis: a review. Pattern Recognit 48(5):1773–1796
18. Liu H, Brown DJ, Coghill GM (2008) Fuzzy qualitative robot kinematics. IEEE Trans Fuzzy Syst 16(3):808–822
19. Liu H, Coghill GM, Barnes DP (2009) Fuzzy qualitative trigonometry. Int J Approx Reason 51(1):71–88
20. Mel BW (1997) Seemore: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. Neural Comput 9(4):777–804
21. Meng Q, Tholley I, Chung PW (2014) Robots learn to dance through interaction with humans. Neural Comput Appl 24(1):117–124
22. Rogez G, Orrite C, Guerrero J, Torr PH (2014) Exploiting projective geometry for view-invariant monocular human motion analysis in man-made environments. Comput Vis Image Underst 120:126–140
23. Rudoy D, Zelnik-Manor L (2012) Viewpoint selection for human actions. Int J Comput Vis 97(3):243–254
24. Shen Q, Leitch R (1993) Fuzzy qualitative simulation. IEEE Trans Syst Man Cybern 23(4):1038–1061
25. Weinland D, Boyer E, Ronfard R (2007) Action recognition from arbitrary views using 3d exemplars. In: International conference on computer vision (ICCV), pp 1–7
26. Weinland D, Ronfard R, Boyer E (2006) Free viewpoint action recognition using motion history volumes. Comput Vis Image Underst 104(2):249–257
27. Weinland D, Ronfard R, Boyer E (2011) A survey of vision-based methods for action representation, segmentation and recognition. Comput Vis Image Underst 115(2):224–241
28. Yilma A, Shah M (2005) Recognizing human actions in videos acquired by uncalibrated moving cameras. In: International conference on computer vision (ICCV), pp 150–157
29. Chan CS, Coghill GM, Liu H (2011) Recent advances in fuzzy qualitative reasoning. Int J Uncertain Fuzziness Knowledge-Based Syst 19(3):417–422