

Fuzzy Action Recognition for Multiple Views within Single Camera

Chern Hong Lim and Chee Seng Chan

Centre of Image and Signal Processing

Faculty of Computer Science and Information Technology

University of Malaya, 50603 Kuala Lumpur, Malaysia

Email: ch_lim@siswa.um.edu.my; cs.chan@um.edu.my

Abstract—To be able to perform human action recognition from multiple views is a great challenge in the field of computer vision. State-of-the-art solutions have been focusing on building a 3D action model from multiple views in a multi, calibrated cameras' environment. Promising results were achieved; however, these approaches tend to assume that human action is performed frontal-parallel to each of the multiple cameras. In a real world scenario, this is not always true and the overlapping regions in such systems are very limited. In this paper, we proposed a fuzzy action recognition framework for multiple views within a single camera. We adopted fuzzy quantity space in the framework and introduced a new concept called the Signature Action Behaviour to model an action from multiple views and represent it as fuzzy descriptor. Then, distance measure is applied to deduce an action. Experimental results showed the efficiency of our proposed framework in modeling the actions from different viewpoints and styles.

I. INTRODUCTION

Human motion analysis (HMA) is one of the very popular researches nowadays due to wide spectrum of applications. The usages of HMA system cross over many domains, for examples, in video surveillance, healthcare, gaming, sports, and military activities. Unfortunately, variation in human appearances, sizes, postures, and individual styles when performing an action makes action recognition a daunting task. Besides that, actions perform from different viewpoints has made the problem harder. Fig. 1 illustrates the variation of motion patterns of an action performed from different viewpoints in a single camera.

Though, most state-of-the-art view-invariant human motion analysis [1]–[5] are found to be deviated from practical solution to serve the purpose. Most of the works assumed that the subjects always perform an action frontal-parallel to the camera; and so built a 3D action model as an action template for matching purpose. This assumption has few limitations. First of all, in a real world environment, subjects are not always frontal-parallel with respect to each of the cameras. Secondly, it is very rare to find a multiple cameras system in public space that has many overlapping regions. Authorities always tend to cover as much areas as possible with a limited number of cameras. The existence of a region that overlaps with multiple cameras, from different viewing angles and the appearance of a subject performing an action that is frontal-parallel to the such cameras are very limited. Therefore, the

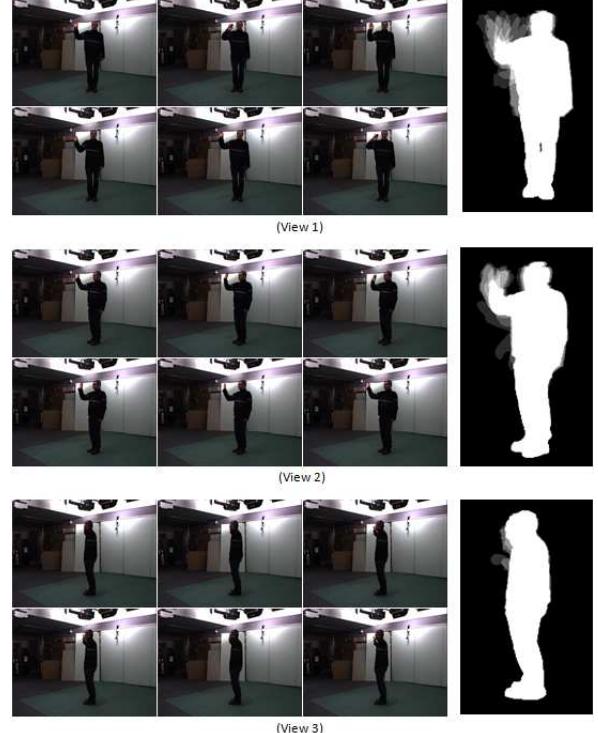


Fig. 1. Motion History Image of ‘wave hand’ action from different viewpoints. It can be noticed that the motion patterns are differed from each other in different subject viewpoint within a single camera.

3D action model that was built based on the above assumption might not work in a real world environment.

With the aforementioned problems, hence, there is a need to have a HMA system for multiple views within a single camera to handle view-invariant action recognition. In this paper, our works focus on HMA system from a single camera and the challenges here is to develop a framework that is able to model and recognize human action performed in different viewpoints and styles, within a single camera. What makes our works different from the others is we intend to model the same action from different viewpoints using the information extracted from the viewpoint itself instead of correlating the information from multiple cameras in a closed controlled environment.

The proposed framework adopted Fuzzy Quantity Space

(FQS) which is built on top of fuzzy set theory [6] and fuzzy qualitative reasoning (FQR) [7], [8] which are able to cope with uncertainties and ambiguities. Readers are encouraged to refer to [9] for the recent advances in fuzzy qualitative reasoning. First of all, we introduced a novel concept named Signature Action Behaviour (SAB) to capture the significant characteristic of an action and apply fuzzy descriptor to represent it. Secondly, for each action from different viewpoints, we built a Fuzzy Descriptor Action Model (FDAM). Finally, distance measure is employed to deduce an action.

In summary, our contributions are: 1) we proposed a fuzzy action recognition framework for multiple views within a single camera. This is in contrast to conventional solutions which focused on multiple views from multiple cameras. 2) we introduced SAB to capture the motion performed in a particular view; and FDAM to model actions from multiple views.

The rest of the paper is organized as follows. Section II covers the related works in HMA. Our proposed framework is discussed in section III which consists of the pipeline to obtain fuzzy descriptor to represent an action. Section IV explains how we build the FDAM for the actions and viewpoints by combining them together in a manifold model. Section V describes how action classification is done by comparing the fuzzy descriptors using distance measure. Section VI discusses the experiments settings and results. Finally, we conclude with suggestions of future work in Section VII.

II. RELATED WORKS

HMA in computer vision has been studied extensively for decades due to the demand and promising growth in high specification of camera technology. Comprehensive surveys of the trends in HMA can be found from [10]–[13] and the survey that closer to our work is [14]. Previous works in HMA can be categorized into two categories which are view-dependent and view-independent approaches.

Most of the works fall into the former categories which assume that all actions are performed with a fix viewpoint [15]–[18]. Notable works in space time approach such as [15] that uses temporal information to build two dimensionary binary of foreground images called Motion-energy image (MEI) and scalar value of foreground images called Motion-history image (MHI) to represent an action, template matching is applied to the pair of MEI and MHI which provides promising results in recognition of a series of ballet actions. [17] proposed an approach which utilizing Poisson solution to estimate the moving torso and protruding limbs that are useful for action recognition, detection and clustering. [16] extended the idea of spatial interest points into spatio-temporal domain where a descriptor is built to classify an event.

With more relate to our works, fuzzy approaches [18], [19] have shown some promising results in HMA. They proposed a tractable and feasible action recognition algorithm which combine fuzzy qualitative robot kinematics [20], [21] with human motion tracking and recognition algorithms. In their works, they adopted FQS [22] to relax the computational

complexities suffering from visual tracking algorithms while preserving the accuracy in action recognition. However, the aforementioned approaches are suffered from the limitation of viewpoint variation. They assumed that the subject performs in a static viewpoint which normally the interesting features can be extracted easily.

While in the latter category, most of the works focus on multiple cameras approaches to achieve view invariant action recognition [1]–[5]. One of the drawbacks of using multi-camera approach is, it is only applicable to closed controlled environment. It may be impractical to deploy in a real world environment. In these approaches, 2D models are extended into 3D models to reconstruct human shape in a volumetric space. For examples, [1] built a 3D representation of human using multiple cameras called voxel person. They extract simple cues from the voxel person and deduce falling activity with fuzzy logic approach. In [4], they propose a new framework that model actions using 3D occupancy grids, built from multiple viewpoints in an exemplar-based HMM. For recognition, the 3D exemplars are used to produce 2D image information for matching purpose. On the other hand, [5] overcomes the viewpoint invariant problem by introducing 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 recognition, they still require information from different viewpoints when an action is performed and these information must be correlated with other cameras from different viewpoints to train and deduce an action, especially the frontal camera provides most information in extracting useful features for an action. In this paper, we remedy these limitations by proposing a framework to model human action from multiple viewpoints and perform recognition within a single camera. The respective models for each action from different viewpoints are built and represented by fuzzy descriptors that will be explained in more details in following sections.

III. OVERALL FRAMEWORK

In this work, we developed a framework that is able to model human action from different viewpoints within a single camera. For ease of understanding, by referring to Fig. 2, we predefine the viewpoints from left to right as ‘horizontal view’, ‘diagonal view’ and ‘vertical view’, respectively; and represented with $v = \{1 \dots V\}$, where in this paper, $V = 3$. These viewpoints will be used throughout the explanation and the experiments in this paper. The overall framework is shown in Fig. 3.

With reference to Fig. 3, given a video sequence, let us denote it as a sequence of frames, $I_{1\dots N}$, where N is the total number of frames and I_n indicate each frame. To begin with, we first apply viewpoint estimation function, E to the video in order to estimate the subject viewing direction, v (Eq. 1).

$$E : I_{1\dots N} \rightarrow v \quad (1)$$

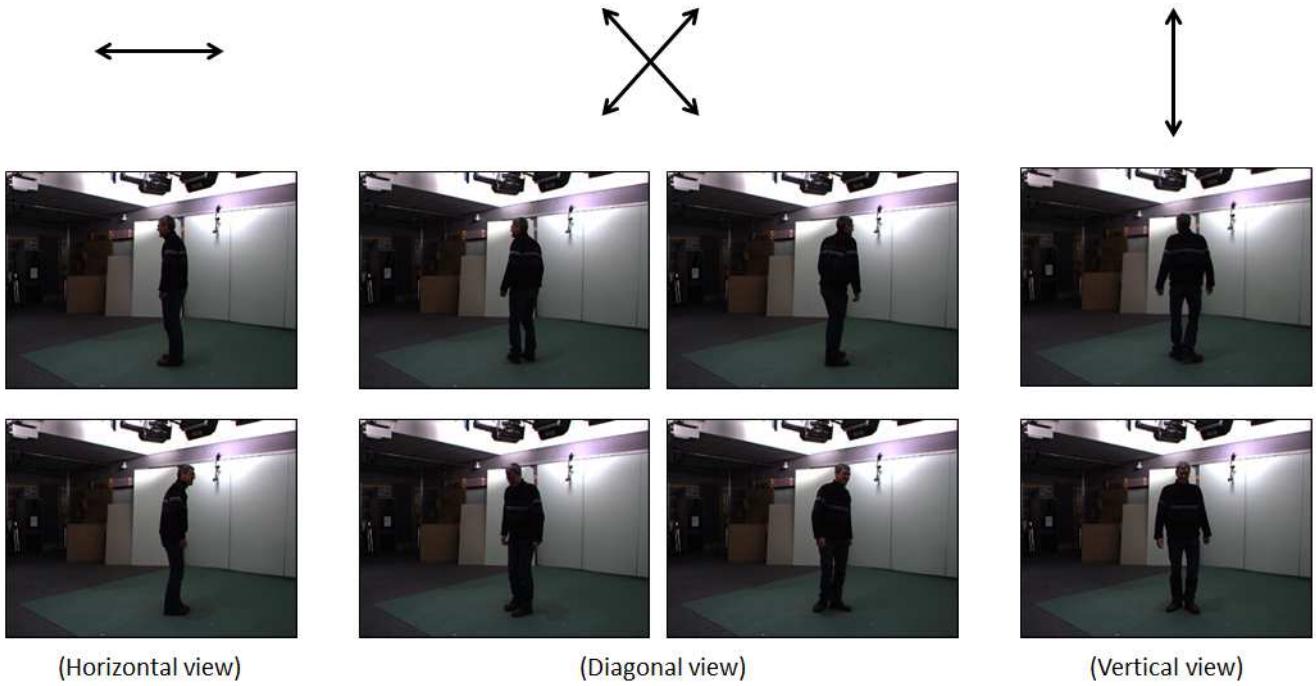


Fig. 2. Definition of viewpoints, from left to right, ‘horizontal view, v_1 ’, ‘diagonal view, v_2 ’ and ‘vertical view, v_3 ’.

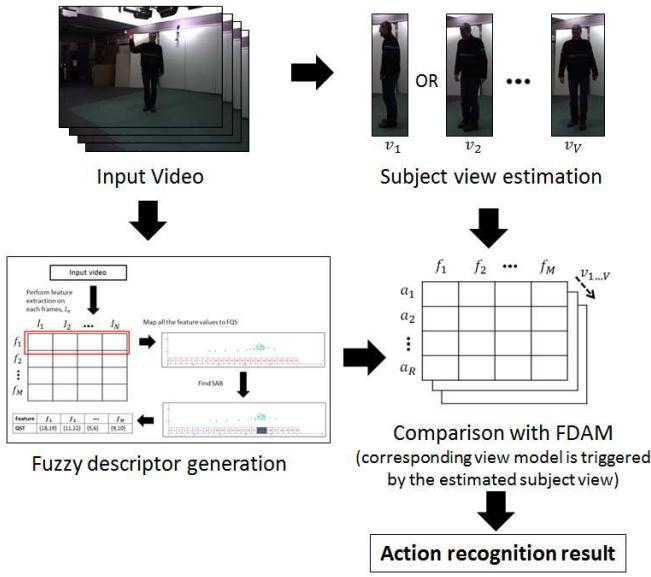


Fig. 3. Overall framework of proposed method.

where $v \in V$ is the output of estimated view that will be used to determine the corresponding FDAM and finally the recognition result is obtained by selecting the shortest distance between the fuzzy descriptor generated from the video and from the corresponding FDAM. However, in this paper, we assumed that prior information of the viewpoints is given; therefore the work is focused on how we build the FDAM that is efficient to model human actions from different styles and viewpoints.

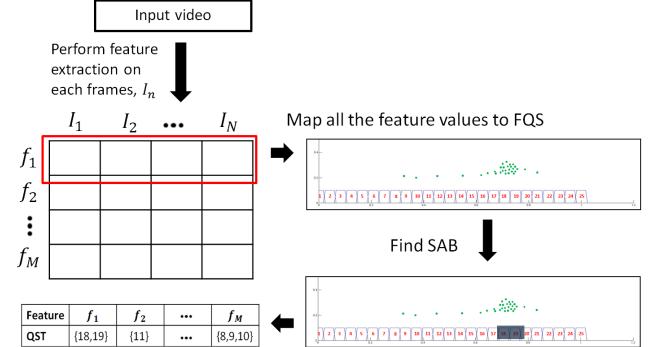


Fig. 4. Fuzzy descriptor generation pipeline.

In order to obtain FDAM, we generate fuzzy descriptor to represent each of the actions performed from each viewpoint (Fig. 4). Our fuzzy descriptor generation process cycle includes feature extraction, FQS mapping of extracted features values, and at last we discover the SAB. Further explanations for each of these steps are in following sections.

A. Feature extraction

Feature extraction is a vital step here where the features are the elements that we wish to model and represent them in a meaningful way that can be used to signify an action. We apply feature extraction function, T on each frame, I_n and the output yields a feature vector for each frame with \mathbb{R}^M where M indicates the total number of features. Putting together all the feature vectors extracted from all the frames, we obtained a matrix of $\mathbb{R}^{M \times N}$ (Eq. 2).

$$T : I_{1 \dots N} \rightarrow \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1N} \\ f_{21} & f_{22} & \cdots & f_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ f_{M1} & f_{M2} & \cdots & f_{MN} \end{bmatrix} \quad (2)$$

In our proposed framework, it accepts any appropriate features, for instance, shape representation, motion flow, texture, etc. provided that they are able to describe an action. In this paper, we capture shapes and motion patterns with the aid of Poisson equation [26] and motion estimation [24]. The output from this step will be mapped into FQS to discover SAB in following sections.

B. Fuzzy Quantity Space Mapping

FQS [22] is generated by a finite discrimination of the underlying range of each variable of a system being modeled. The FQS will have the desirable properties of finiteness and coverage, as long as the system contains a finite number of variables. Granularity in the FQS is obtained by the arbitrariness of the discrimination of the numeric ranges of system variables that are assumed to be of interest. Hence, a subset of a numeric range can be translated to one qualitative value according to what is needed in a particular modeling process. The adoption of fuzzy subsets has a direct distinct advantage over the traditional crisp representations when considering granularity.

In fact, if one intends to describe the qualitative values of system variables only in terms of the crisp subsets of the underlying real range of the variables, the mapping from the real range to a quantity space will result in the search for the limits of the real numbers served as the boundaries between (dis-jointly) adjacent qualitative values within the quantity space. This usually incurs severe difficulties in determining these limits [8]. The fuzzy representation of qualitative values is more general than ordinary (crisp) interval representations, since it can represent not only the information stated by a well-determined real interval but also the knowledge embedded in the soft boundaries of the interval. Thus, FQS removes, or largely weakens (if not completely resolving), the boundary interpretation problem, achieved through the description of a gradual rather than an abrupt change in the degree of membership of which a physical quantity is mapped onto a particular qualitative value. It is, therefore, closer to the common sense intuition of the description of a qualitative value.

The construction of FQS utilizes a parametric approximation of the membership function where the membership distribution of a normal convex fuzzy number is approximated by the 4-tuple, $[a, b, \alpha, \beta]$ [22]. In this paper, we represent the x -axis of the FQS as in [23] with Eq. 3 and it is normalized to be in the range of $[0 \ 1]$. This means that the features that we extracted must be also normalized before we can map them into FQS.

$$FQS^x = \{QST_x^1, QST_x^2, \dots, QST_x^J\} \quad (3)$$

where QST_x are the fuzzy qualitative states [23] along the x -axis, and J is the total number of QST_x . From the previous step, we obtained the feature matrix. In this step, we map the normalized values of each feature, f_m across the frames, $1 \dots N$, into FQS (Eq. 4) as shown in Fig. 5 (a simulated image described the mapping of feature values into a FQS).

$$f_{m(1 \dots N)} \mapsto FQS_{f_m}^x \quad (4)$$

after the mapping is done, we are able to perform frequency counting of the number of data that falls into each fuzzy qualitative state denoted as $q(QST_x)$ and the outcome of the mapping of each feature in FQS is designated as Eq. 5.

$$FQS_{f_m} = \{q(QST_x^1), q(QST_x^2), \dots, q(QST_x^J)\} \quad (5)$$

The intention of using FQS in modeling feature values for an action is because with fuzzy tuple, we can model the uncertainties within a tolerance range for the feature values and at the same time we are able to retrieve the information about which QST is involved. This information will be used in next section to discover the SAB.

C. Discover Signature Action Behaviour

When human performing an action, we infer an action by observing the characteristic of a series of motion within a short period. Within this moment, we will conclude an action by understanding the behaviour of the motions. For example, if a person is repeatedly raising and put down the hand, we understand that it is a waving action. Where in here, raising and put down motion we consider as signature behaviour for the waving action. In computational aspect, using the same example, if a motion detection algorithm is able to provide us the measurement of these information over a normalized scale, we consider that the prominent density on this scale are assumed to be the signature behaviour of the particular action.

This is what inspired us to propose the concept of capturing SAB to represent an action. With continue from the previous step, we compute the weight for each QST_x , denoted as W_j by normalize all the $q(QST_x)$ with the maximum $q(QST_x)$ as in Eq. 6.

$$W_j = \frac{q(QST_x^j)}{\max(q(QST_x))} \quad (6)$$

after that, SAB is discovered by finding the QST_x (s) which own the W that is $> Thres$, where $Thres$ represent the threshold defined by user and in this work, we use $Thres = 0.7$. We designated SAB for each action as D_{f_m} as in Eq. 7. From here, one can notice that the element of D_{f_m} can be more than one, provided the W value of the QST is more than the threshold. The intention is we do not ignore any potential QST that can be SAB.

$$D_{f_m} = QST_x(W > Thres) \quad (7)$$

For example from Fig. 5, the SAB for that particular feature most likely fall into QST_x^{18} , and QST_x^{19} as both

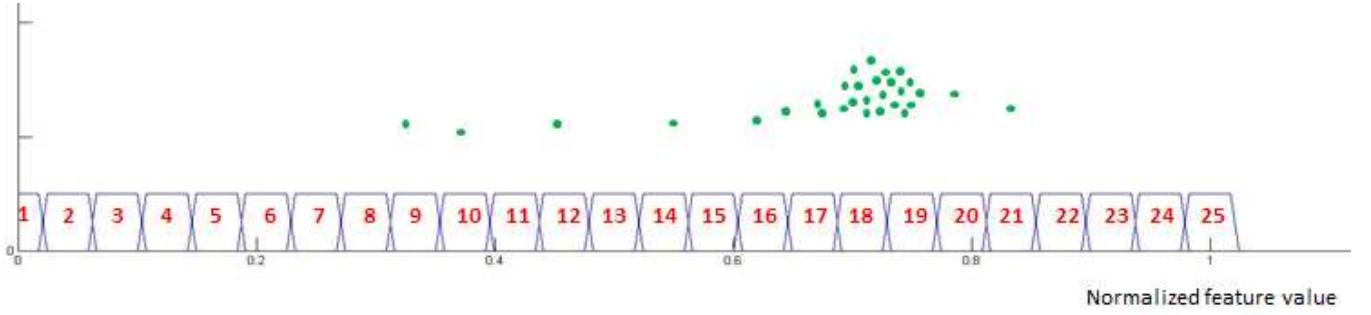


Fig. 5. Simulation of feature values mapping into FQS with number of QST_x , $J = 25$. From here, we observe that the state QST_x^{18} and QST_x^{19} possess highest density among the other states.

fuzzy qualitative states possess high density compare to others. Conceptually SAB is a significant representation of an action performs by a subject. Ideally if this is compared with another subject who is also performing similar action, the similarity will be higher than compared to the other actions.

D. Building Fuzzy Descriptor Model

The significant differences between fuzzy descriptor, FD compared to other conventional descriptors are, fuzzy descriptor allows the occupancy of **multi-instance** SAB for each feature. This mean that the element of D_{f_m} can be more than one as stated above. In order to form a complete descriptor, we repeatedly discover the SAB for the rest of the features and concatenate them together as Eq. 8.

$$FD = \{D_{f_1}, D_{f_2}, \dots, D_{f_M}\} \quad (8)$$

Fuzzy descriptors are used to compare actions between the testing video with the corresponding FDAM determined by the estimated viewpoint. By measuring the distances between them, the shortest distance is deduced as the most likely action that this subject is performing.

IV. BUILDING ACTION MODEL

In this section, we intend to build FDAM that significantly represents the actions from different viewpoints and at the same time manage to cope with style invariant issue. Similar to other supervised methods, in order to achieve that, we need to have training samples. The training samples are represented by human as subject, s_o where $o = 1, 2, \dots, O$ that performs an action a_r where $r = 1, 2, \dots, R$ from view, v_1, v_2, \dots, v_V . With reference to Fig. 6, we apply previous steps to all the training subjects for all the actions and viewpoints to generate the respective fuzzy descriptors and then combine them into a manifold model, which we called it as FDAM (ref to Fig. 6).

In nature, different human tends to perform an action with varieties of styles. The difference of styles including physical differences (such as human appearances, sizes, postures, etc.) and dynamic differences (speed, motion pattern, etc.). In order to model these variations, our framework constructs a fuzzy descriptor to represent an action that is tolerance with different styles. The adoption of FQS provides the capability to cover

and model the small differences in the features values extracted from an action and this is to cope with dynamic differences in action.

As aforementioned, we tend to obtain variation of QST_x as SAB in respective D_{f_m} due to different styles. For instance, different QST_x we might obtained from people who performed waving with right hand and with left hand (physical difference). Herein, we use Union aggregation to fuse them together (Eq. 9), as in heuristic understanding, both are performing the same action.

$$DU_{f_m} = \bigcup_{o=1}^O \{D_{f_m}\}_{s_o} \quad (9)$$

The same goes to the SAB for other features. Finally, we concatenate all these DU_{f_m} to construct the fuzzy descriptor for an action at the particular view and represent it as (Eq. 10).

$$FDAM_{a_r v} = \{DU_{f_1}, DU_{f_2}, \dots, DU_{f_M}\}_{a_r v} \quad (10)$$

Finally, we obtain the FDAM for different actions and from different viewpoints.

V. ACTION RECOGNITION

In this section, we are going to explain how to perform action recognition from the FDAM. First we need to estimate the view, v of the action (Eq. 1). The estimated v is used to determine the FDAM that will be used in the inference step, then distance measure is applied to deduce an action. Let us denote the testing sample as G while the action model as H .

With fuzzy descriptor, ordinary distance measure may not work if the SAB element is more than one. Instead, we use Eq. 11 to obtain the minimum L2-norm of both SAB from testing sample, D^G and the action model, DU^H .

$$d(D^G, DU^H) = \min_{x \in D^G, y \in DU^H} [d(x, y)] \quad (11)$$

To measure the distance between the fuzzy descriptor of the training sample, FD^G and action model, $FDAM_{a_r v}^H$ (assume that v is known), we sum all the L2-norm of SAB as Eq. 12.

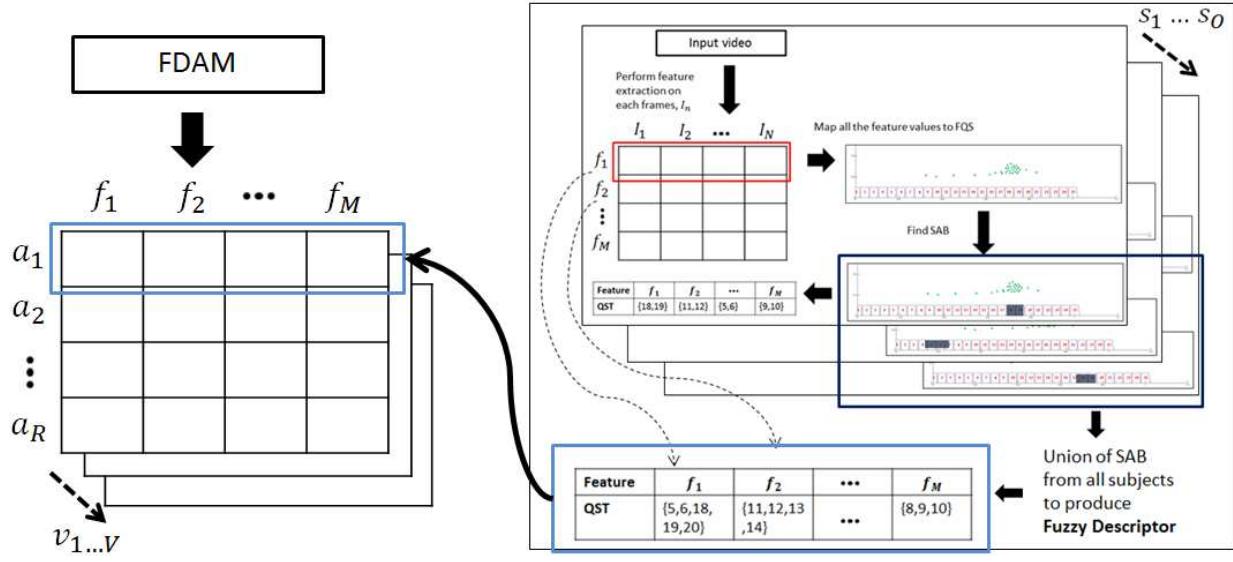


Fig. 6. Building of FDAM by constructing the fuzzy descriptor for each action a across the subject s_1 to s_O . The process is repeated to obtain the action model for $v_1 \dots v$.

$$dist(FD^G, FDAM_{a,r}^H) = \sum_{m=1}^M d(D_{f_m}^G, DU_{f_m}^H) \quad (12)$$

The final result is inferred by finding the minimum distance between the fuzzy descriptor of the actions (Eq. 13).

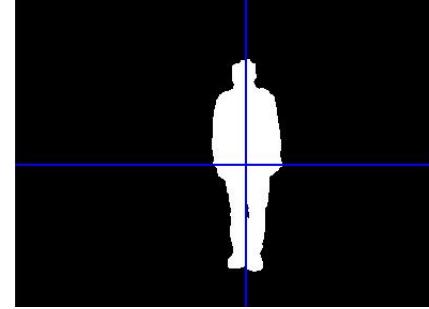
$$result = \forall_{a_r \in R} \min[dist(FD^G, FDAM_{a,r}^H)] \quad (13)$$

The *result* indicates the particular fuzzy descriptor to be the nearest toward fuzzy descriptor for the respective action in FDAM with shortest distance. And thus, the system output it as the result of most likely action performs by the testing subject.

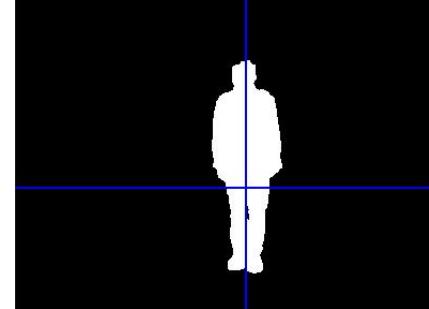
VI. EXPERIMENT

In order to test the efficiency of our proposed framework, we employ IXMAS action dataset which is available publicly from (<http://4drepository.inrialpes.fr/public/viewgroup/6>) in our experiments. From the dataset, we selected 11 actors, 5 actions ($a_1 = \text{scratch head}$, $a_2 = \text{sit down}$, $a_3 = \text{get up}$, $a_4 = \text{wave hand}$, and $a_5 = \text{kick}$), and the 3 subject viewpoints as defined in Fig. 2. As for the FQS, we have chosen 25 tuples ($J = 25$) to represent the feature values. In the pre-processing step, we obtain the silhouette image of the subject. According to [25], silhouette image is sufficient to capture the activity describe by a frame sequence, and at the same time conserve the computational time. In feature extraction, we extract the shape features and the motion patterns of different parts of the limbs of a body but not limited to. As a reminder, reader can use other appropriate features to incorporate with our framework.

Adopting the idea from [24], we segment the human body into four parts. From there, we determine the motion of right



(a) [24]'s method



(b) Our method

Fig. 7. Comparison of the methods used to perform segmentation of body into four parts.

hand (Top Left segment), left hand (Top Right segment), right leg (Bottom Left segment) and left leg (Bottom Right segment). However, in [24], they used an inappropriate method to determine the middle point to make the segment. This could end up with inappropriate estimation of the body segment. In this paper, we adopted Poisson solution [26] to estimate the lower part of the torso of a human body and use it to segment

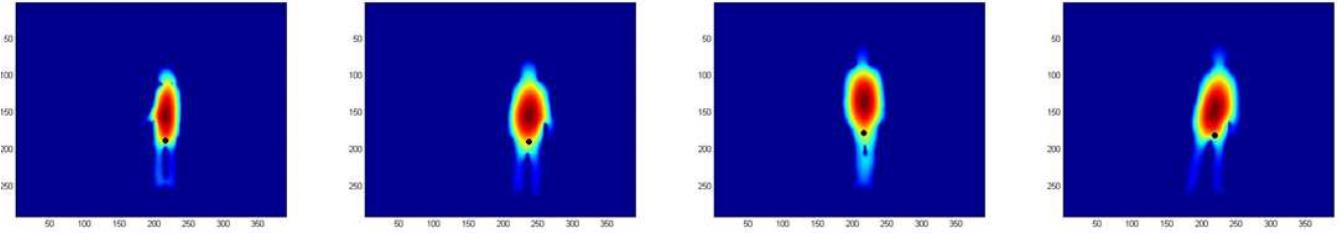


Fig. 8. The middle lower part of torso estimated using Poisson equation (denoted as a black dot). It is proved that this technique works on human with variation of sizes, heights, and postures.

the human body into four parts. We anticipated that this will be able to handle different human anatomy, e.g. body sizes, heights, and postures) as shown in Fig. 8.

Fig. 7 shows the comparison of body segmentation using the method from [24] and our method. We observed in Fig. 7(a) that body with portions of the hand falls into Bottom Left segment and Bottom Right segment which in ideal case, only legs are required in these segments. While on the other hand, Fig. 7(b) shows our method estimates the body segments correctly. Then, a feature vector is extracted for each frame. In classification, we applied ‘leave-one-out’ approach to test the efficiency of our framework.

A. The accuracy of action recognition

Fig. 9 shows the average accuracy of the recognition of five actions and from three viewpoints.

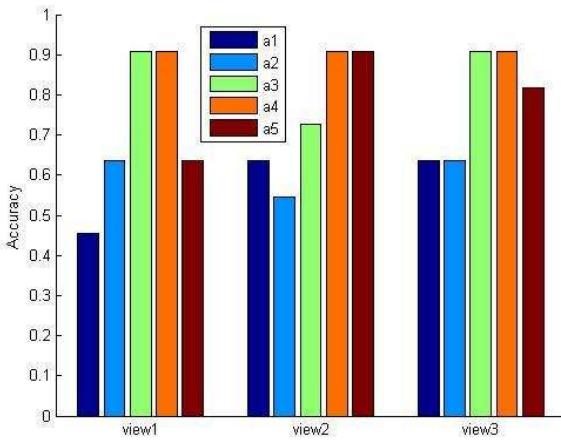


Fig. 9. Bar chart of recognition accuracy for action a_1, a_2, a_3, a_4 , and a_5 perform with view1, v_1 , view2, v_2 , and view3, v_3 .

From the results, we observe that action such as ‘get up’, ‘wave hand’ and ‘kick’ are much consistent and have better recognition rate; while the ‘scratch head’ and ‘sit down’ actions obtain lower accuracy. We anticipate that these actions might confuse with other actions which are similar enough to deduce a false positive results. These will be explained in more detail in next section where we show the confusion matrix of recognition for actions from different viewpoints.

B. Confusion matrix for action recognition

Here, in more detailed comparison, we show the confusion matrix of action recognition results for different viewpoints individually. However, in this comparison, sum of each row may be more than one because there might be repeated results on different action, for example, action a_1 can be detected as a_1 and also a_5 . However, these are just minor cases.

TABLE I
CONFUSION MATRIX OF RECOGNITION RATE FOR v_1

	a_1	a_2	a_3	a_4	a_5
a_1	0.4545	0	0	0.8182	0.0909
a_2	0	0.6364	0.0909	0	0.6364
a_3	0	0	0.9091	0	0.0909
a_4	0.3636	0	0	0.9091	0
a_5	0	0.4545	0.0909	0	0.6364

TABLE II
CONFUSION MATRIX OF RECOGNITION RATE FOR v_2

	a_1	a_2	a_3	a_4	a_5
a_1	0.6364	0	0	0.4545	0.0909
a_2	0	0.5455	0	0.1818	0.4545
a_3	0	0	0.7273	0.1818	0.0909
a_4	0.1818	0	0	0.9091	0
a_5	0	0	0	0.0909	0.9091

TABLE III
CONFUSION MATRIX OF RECOGNITION RATE FOR v_3

	a_1	a_2	a_3	a_4	a_5
a_1	0.6364	0	0.1818	0.2727	0.0909
a_2	0.0909	0.6364	0	0	0.2727
a_3	0	0	0.9091	0	0.0909
a_4	0.2727	0	0	0.9091	0.0909
a_5	0.1818	0.0909	0	0.0909	0.8182

From the confusion matrices, we notice that action a_1 which is ‘scratch head’ is always confused with action a_4 which is ‘wave hand’. We know that, these two actions are very similar in terms of the motion pattern. Especially in view $v = 1$, it is very difficult to differentiate these two actions as they were preformed from the side. With reference to Fig. 10, this is reasonable as in real environment, human have difficulty to see the action from this viewing direction too.

On the other hand, the best recognition rates are from view $v = 3$ because all actions are performed facing the camera



(a) Scratch head



(b) Wave hand

Fig. 10. Example of frames for ‘scratch head’ and ‘wave hand’ action.

and therefore, our proposed framework is able to model the action efficiently. Fortunately, the results have proven that our proposed framework owns the potential to model and infer the actions that were performed with different styles and subject views.

VII. CONCLUSION

In this paper, we presented a framework to model the actions performed from different viewpoints within single camera. The actions are represented with fuzzy descriptor which is built on top of the novel concept called SAB. The adoption of FQS in our framework endowed the ability to model the actions with variation of styles. Finally, we build FDAM which consists of the action models from different viewpoints. In action classification, fuzzy descriptors are used to measure the distance between the testing subject with the FDAM by using the distance measure explained in the context. Shortest distance determines the final output of the recognition result. The experiments showed the efficiency of our framework to model and recognize the action of different styles and different viewpoints in a single camera. In future, we are going to investigate the relationship and correlation between the fuzzy descriptors of the viewpoints and build a manifold descriptor to indicate an action instead of having separate view models for every action.

REFERENCES

- [1] D. Anderson, R. Luke, J. Keller, M. Skubic, M. Rantz, and M. Aud, “Modeling human activity from voxel person using fuzzy logic,” *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 1, pp. 39–49, feb. 2009.
- [2] M. Ahmad and S.-W. Lee, “Hmm-based human action recognition using multiview image sequences,” in *International Conference on Pattern Recognition*, vol. 1, 2006, pp. 263–266.
- [3] D. Weinland, R. Ronfard, and E. Boyer, “Free viewpoint action recognition using motion history volumes,” *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 249–257, 2006.
- [4] D. Weinland, E. Boyer, and R. Ronfard, “Action recognition from arbitrary views using 3d exemplars,” in *International Conference on Computer Vision*, oct. 2007, pp. 1–7.
- [5] A. Yilmaz and M. Shah, “Recognizing human actions in videos acquired by uncalibrated moving cameras,” in *International Conference on Computer Vision*, vol. 1, oct. 2005, pp. 150–157 Vol. 1.
- [6] L. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [7] D. Dubois and H. Prade, “Fuzzy arithmetic in qualitative reasoning,” *Lecture Notes in Control and Information Sciences*, vol. 212, pp. 457–467, 1989.
- [8] Q. Shen and R. Leitch, “Fuzzy qualitative simulation,” *IEEE Transactions on Systems, Man And Cybernetics*, vol. 23, no. 4, pp. 1038–1061, 1993.
- [9] C. S. Chan, G. M. Coghill, and H. Liu, “Recent advances in fuzzy qualitative reasoning,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 19, no. 3, pp. 417–422, 2011.
- [10] L. Wang, W. Hu, and T. Tan, “Recent developments in human motion analysis,” *Pattern Recognition*, vol. 36, no. 3, pp. 585–601, 2003.
- [11] T. B. Moeslund, A. Hilton, and V. Kruger, “A survey of advances in vision-based human motion capture and analysis,” *Computer Vision and Image Understanding*, vol. 104, pp. 90–126, 2006.
- [12] R. Poppe, “A survey on vision-based human action recognition,” *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, 2010.
- [13] J. Aggarwal and M. Ryoo, “Human activity analysis: A review,” *ACM Comput. Surv.*, vol. 43, no. 3, pp. 16:1–16:43, Apr. 2011.
- [14] X. Ji and H. Liu, “Advances in view-invariant human motion analysis: A review,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 1, pp. 13–24, jan. 2010.
- [15] A. Bobick and J. Davis, “The recognition of human movement using temporal templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, mar 2001.
- [16] I. Laptev, “On space-time interest points,” *International Journal of Computer Vision*, vol. 64, pp. 107–123, 2005.
- [17] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, “Actions as space-time shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2247–2253, dec. 2007.
- [18] C. S. Chan and H. Liu, “Fuzzy qualitative human motion analysis,” *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 851–862, 2009.
- [19] C. S. Chan, H. Liu, D. J. Brown, and N. Kubota, “A fuzzy qualitative approach to human motion recognition,” in *International Conference on Fuzzy Systems*, 2008, pp. 1242–1249.
- [20] H. Liu, D. J. Brown, and G. M. Coghill, “Fuzzy qualitative robot kinematics,” *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 3, pp. 808–822, 2008.
- [21] H. Liu, “A fuzzy qualitative framework for connecting robot qualitative and quantitative representations,” *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1522–1530, dec. 2008.
- [22] H. Liu, G. M. Coghill, and D. P. Barnes, “Fuzzy qualitative trigonometry,” *International Journal of Approximate Reasoning*, vol. 51, no. 1, pp. 71–88, 2009.
- [23] C. H. Lim and C. S. Chan, “A fuzzy qualitative approach for scene classification,” in *International Conference on Fuzzy Systems*, 2012, pp. 1–8.
- [24] C. H. Lim, A. Kadyrov, C. S. Chan, and H. Liu, “Motion detection based on simulated depth measurement,” IOS Press, 2012, p. 335.
- [25] M. Lewandowski, D. Makris, and J.-C. Nebel, “View and style-independent action manifolds for human activity recognition,” in *European Conference on Computer Vision*. Springer Berlin Heidelberg, 2010, pp. 547–560.
- [26] L. Gorelick, M. Galun, E. Sharon, R. Basri, and A. Brandt, “Shape representation and classification using the poisson equation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1991–2005, dec. 2006.