Snippet Based Trajectory Statistics Histograms for Assistive Technologies

Ahmet Iscen^{1,3}, Yijie Wang^{2}, Pinar Duygulu^{1}, and Alex Hauptmann^{2}

^{1} Bilkent University
^{2} Carnegie Mellon University
^{3} Inria Rennes

Abstract. Due to increasing hospital costs and traveling time, more and more patients decide to use medical devices at home without traveling to the hospital. However, these devices are not always very straight-forward for usage, and the recent reports show that there are many injuries and even deaths caused by the wrong use of these devices. Since human supervision during every usage is impractical, there is a need for computer vision systems that would recognize actions and detect if the patient has done something wrong. In this paper, we propose to use Snippet Based Trajectory Statistics Histograms descriptor to recognize actions in two medical device usage problems; inhaler device usage and infusion pump usage. Snippet Based Trajectory Statistics Histograms encodes the motion and position statistics of densely extracted trajectories from a video. Our experiments show that by using Snippet Based Trajectory Statistics Histograms technique, we improve the overall performance for both tasks. Additionally, this method does not require heavy computation, and is suitable for real-time systems.

Keywords: Assisted living systems, medical device usage, action recognition

1 Introduction

Understanding, predicting, and analyzing people's actions can be a challenging task not only for computers, but also for humans. Even though there have been many breakthrough work in the domain of human action recognition, a significant amount of this research is about recognizing ordinary actions from movies [13] or sports [19]. Most of these solutions can be extended to introduce solutions that may have a more direct impact in our lives. Some of these applications can include video surveillance, anomaly detection, and daily activity analysis.

In this paper, we focus on assistive systems on home medical device usage, more specifically, infusion pump and inhaler usage. Infusion pumps are used to deliver fluid and medication into a patient's body, and inhaler devices are used for asthma therapy. Due to increasing costs, more and more patients decide to use these devices at home to save money and time. However, with lack of experience and failure to follow instructions, personal usage of these devices may be ineffective, or even worse, cause fatal problems. In fact, U.S. Food and Drug Administration reported 56,000 incidents between 2005 and 2009, including deaths and injuries, caused by infusion pump usage [8].

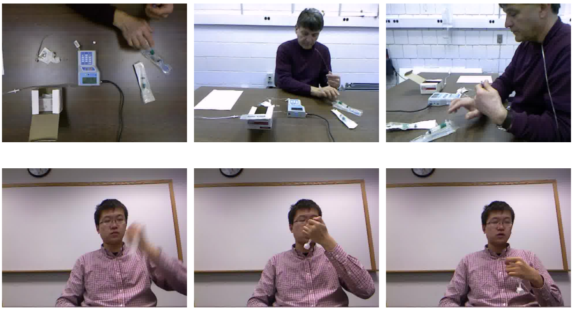


Fig. 1. Our task is to detect activities for medical device usage, more specifically for infusion pump and inhaler devices. An example of infusion pump usage is shown on the first row, and inhaler usage is shown on the second row. There are recordings from 3 different camera angles for infusion pump usage.

With the advancements in computer vision, and impracticality of having a supervisor controlling the correct medical device usage of the patient every time they need to use it, there is a clear need for assistive systems to detect flaws and mistakes during the usage, and report it to the patient. Nevertheless, compared to a typical video analysis system, there may be some limitations of such system. One example is that it must be a real-time system which can detect actions on the fly. Another challenge is that, some actions may only consist of little movements by certain part of the body. This is especially true for infusion pump usage.

On contrary, as such methods are developed for a specific purpose, we can make some assumptions when developing a new method. As an example, we can assume that the camera will usually record the patient from a certain perspective, therefore certain actions may happen in specific positions. In fact, Cai *et al.* make use of this information to detect actions for infusion pump usage [2]. Another assumption we can make is that there most likely will not be a camera motion when the patient performs an action, as the patient is always expected to be seated when the usage takes place.

In this paper, we use Snippet Based Trajectory Statistics Histograms technique [10] to detect actions for assistive system applications. This descriptor encodes the position and movement statistics of a motion, which makes it a suitable candidate for this application due to the aforementioned assumptions. Also, because of their simplicity, it is possible to implement and use them in real time applications. Note that the proposed method only needs RGB videos rather than the depth information which is regularly used for such applications [2, 29]. We believe that, although the depth information of a video is extremely useful for detecting some actions, an ordinary patient may not always have an access to a device with such capability. Therefore, we propose to investigate a method which would only use RGB video, and no depth information at all.

2 Related Work

Activity recognition has always been a well studied topic in computer vision. Even though most of the work focuses on finding "ordinary" actions [1], applying them to the medical device usage or other assistive living applications may not always be straightforward. For example, dense trajectories method [27] by Wang *et al.* works very well for recognizing complex activities, however in this work we show that the simpler Snippet Based Trajectory Statistics Histograms method outperforms it for the medical device usage applications.

Some of the research for activity recognition focuses on applications involving daily activities. One of these applications is anomaly detection, which can be used for surveillance or assistive technologies. Roshtkari *et al.* [22] detects anomalous behaviors by utilising a hierarchical codebook of dense spatio-temporal video volumes. Unusual events are formulated as a sparse coding problem in [33]. Another interesting direction about unusuality detection is learning the crowd behavior [20], and then detecting anomaly as outliers [24]. Ziebart *et al.* predicts people's future locations [34] and Kitani *et al.* [12] forecasts human actions by considering the physical environment. Other works involving daily activities include daily action classification or summarization by egocentric videos [7, 17, 14], fall detection [15], and classification of cooking actions [26, 23, 21, 11].

Most of the research in assisted living domain requires the use of many different sensors [4, 28, 31, 18]. This kind of setup is usually very costly and not practical for the user. There are also other works that only focus on the use of videos [3]. Fleck and Strasser [9] use cheap smartphone cameras for assisted living monitoring system. Smartphones are also used for obstacle detection for visually impaired people [25, 16]. Also, some applications help visually impaired people with their grocery shopping [30], and sonify images for them [32].

Our work is most similar to [2, 29], who also analyzes the use of infusion pump and inhaler devices. One of the main differences is that we develop a method only using RGB videos, and not the depth information. Also, unlike [2] we aim for a real-time application which would be able to detect actions rapidly and warn the user if needed.

3 Method

When analyzing the medical device usage, there are two properties that we can make use of; the magnitude of motion and position. Since we can assume that there will not be any camera motion, we can consider all the motion information to be relevant, and do not have to take into account any noisy motion. Also, by assuming that the position of the user will be more or less the same every time, we try to exploit the information about where each action takes place.

3.1 Snippet Based Trajectory Statistics Histograms

Originally introduced to detect unusual actions in videos [10], Snippet Based Trajectory Statistics Histograms method encodes the position and motion information in small intervals. The visualization of the algorithm is shown on Figure 2. It is built on dense sampling technique introduced by Wang *et al.* [27], which is used to produce the original motion trajectories from the video. These trajectories are extracted densely on the grid, and their motion is tracked for every D frames.

Unlike the *trajectory feature* described in [27], which encodes the shape of each trajectory, Snippet Based Trajectory Statistics Histograms encodes the motion information simply by considering their statistical information. That is, for each trajectory T, we only use its length (l), variance along x-axis(v_x), and y-axis (v_y). Length information is used to distinguish between fast and small motions. Since trajectories are tracked for a fixed number of frames, the longer ones correspond to faster motions, whereas the smaller ones correspond to motions without much movement. Additionally, in order to encode the direction and the spatial extension of the motion, variance of motion in x and y coordinates are also used.

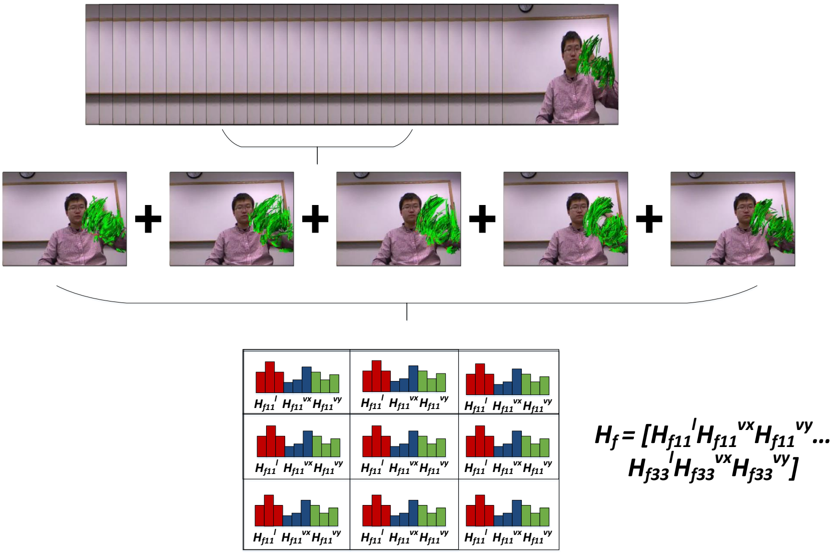

+
+
+
+
H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy}
H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy}
H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy}
H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy}
H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy}
H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy}
H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy}
H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy}
H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy}
$H_f = [H_{f11}^{l} H_{f11}^{vx} H_{f11}^{vy} \ldots$
$H_{f33}^{l} H_{f33}^{vx} H_{f33}^{vy}]$

Fig. 2. Visualization of Snippet Based Trajectory Statistics Histograms technique. A small interval of a video is taken, and the motion information is aggregated. A histogram based on the length and the variance statistics of this motion, as well as its spatial location, is created.

In order to aggregate motion statistics along with position, Snippet Based Trajectory Statistics Histograms algorithm finds the average position of each trajectory during D frames which it was tracked for. First, for each trajectory T, we find its average position in x and y coordinates, m_x and m_y respectively, along with its motion statistics. This can be done as follows:

$$
\begin{aligned}
m_x &= \frac{1}{D}\sum_{t}^{t+D-1} x_t, v_x = \frac{1}{D}\sum_{t}^{t+D-1}(x_t - m_x)^2 \\
m_y &= \frac{1}{D}\sum_{t}^{t+D-1} y_t, v_y = \frac{1}{D}\sum_{t}^{t+D-1}(y_t - m_y)^2, \\
l &= \sum_{t}^{t+D-1} \sqrt{(x_{t+1} - x_t)^2 + (y_{t+1} - y_t)^2}
\end{aligned}
\tag{1}
$$

where l is the length, v_x and v_y are the variance of the trajectory in x and y axes respectively.

Now that we have the motion statistics of trajectories, they need to be aggregated in order to describe the overall motion in a small interval of the video. This is done by describing all the motion and position statistics in each small interval, which is also referred as *snippets*.

Snippet Based Trajectory Statistics Histograms are calculated for each frame in the video, and they encode the motion before and after the frame. As an example, for a given snippet size of S at frame f, all the trajectories that end at frame t are aggregated, such that $f - \|S\|/2 \leq t \leq f + \|S\|/2$.

To embed the spatial information of the movement, we divide each frame into $N \times N$ spatial grids. Recall that we have already computed the average x and y positions m_x and m_y for each trajectory. We assign each trajectory to one of the spatial grids based on their m_x and m_y positions, and for each spatial grid, we create 3 histograms by separately quantizing the l , v_x and v_y values of all the trajectories belonging to that grid into b bins. For example, let $H_l(t)$ be the frame histogram of length (l) values of all the trajectories which are tracked until frame t. We create this histogram by concatenating b-bin histograms from each spatial bin, such that:

$$H_l(t) = (H_l(t)_{[1,1]}, \ldots H_l(t)_{[1,N]}, \ldots H_l(t)_{[N,N]}) \tag{2}$$

where $H_l(t)_{[i,j]}, 0 \leq i, j \leq N$ contains the b-bin histogram obtained by trajectories whose average position lies in the $[i,j]^{th}$ cell. The same procedure is also done for the v_x and v_y values of trajectories.

Finally, in order to encode motion and position information for a $snippet$ of size S, and create the Snippet Based Trajectory Statistics Histograms, we sum all the frame histograms neighboring the frame f for each statistic separately:

$$
\begin{gathered}
H_f^l = \sum_{t=f-(\|S\|/2)}^{f+(\|S\|/2)} H_l(t) \quad , \quad H_f^{vx} = \sum_{t=f-(\|S\|/2)}^{f+(\|S\|/2)} H_{vx}(t) \\
H_f^{vy} = \sum_{t=f-(\|S\|/2)}^{f+(\|S\|/2)} H_{vy}(t))
\end{gathered}
\tag{3}
$$

and concatenate them in the end to obtain a $3 \times b \times N \times N$ dimensional histogram for each frame f:

$$H_f = [H_f^l, H_f^{vx}, H_f^{vy}] \tag{4}$$

3.2 Why Snippet Based Trajectory Statistics Histograms

Our goal is to detect actions for medical device usage. We propose to use Snippet Based Trajectory Statistics Histograms method for this domain for a few properties which we believe would be handled by Snippet Based Trajectory Statistics Histograms.

First of all, we know that the camera will be standing still and looking over the patient from a certain angle during the usage. Therefore, certain actions will always occur in certain spatial positions of the frame, since the position of each user will more or less be the same. Because Snippet Based Trajectory Statistics Histograms divides the spatial coordinate into grids, and creates a separate histogram for each grid, we believe that it will help us to distinguish similar actions taking place in different positions.

Secondly, medical device usage may involve very fast actions, such as *shaking* and very slow actions, such as *opening a cap*. By exploiting the statistical information of trajectories, such as the length and the variance of the trajectory motion, Snippet Based Trajectory Statistics Histograms would enable the system to distinguish fast actions from slow actions.

Lastly, due to the simplicity and easiness of the implementation, Snippet Based Trajectory Statistics Histograms can be suitable for real-time applications. In fact, the major overhead of execution is to generate initial trajectories using the dense trajectory method [27], but this can be sped up easily by downsampling the sizes of frames. Once we have the initial trajectories, Snippet Based Trajectory Statistics Histograms requires only a few fast calculations.

4 Experiments

We conduct our experiments for two different scenarios, inhaler device and infusion pump device usage. In this section, we first describe the datasets we used for each task, and explain our results.

4.1 Implementation Details

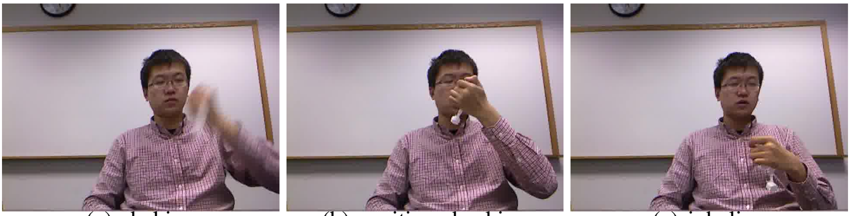
For our implementation of Snippet Based Trajectory Statistics Histograms for medical device usage, we extract initial snippets for $D = 15$ frames. Then, we find Snippet Based Trajectory Statistics Histograms in $N \times N$ spatial grid where $N = 3$, and quantize length and variance information into $b = 5$ bins. Therefore, we have $3 \times 5 \times 3 \times 3 = 135$ dimensional descriptors in the end. The size of each snippet is $S = 10$ frames.

4.2 Baseline

We compare Snippet Based Trajectory Statistics Histograms with other trajectory-based descriptors, such as Trajectory Shape Descriptor [27], histograms of oriented gradients (HOG) [5], histograms of optical flow (HOF) [13], and motion boundary histograms (MBH) [6]. We quantize the other descriptors with bag-of-words using a dictionary of size $k = 100$.

4.3 Inhaler Usage

Dataset Inhaler Dataset [29] has 77 RGB and depth video recordings of users using GlaxoSmithKline Inhalation Aerosol device. The user performs inhaler usage operation while sitting 50-70 cm in front of the camera. The dataset has 4 different action classes, *shaking*, where the user shakes the inhaler device, *position checking*, where the user puts the device about 2 inches in front of his or her mouth, *inhaling* and *exhaling*. However, since *inhaling* and *exhaling* actions cannot be detected visually, and need additional features such as audio, we do not evaluate them. We use recall, precision, and F-score measures for the evaluation of performance in this dataset.



(a) shaking (b) position checking (c) inhaling

Fig. 3. Some of the actions in the Inhaler Dataset. Although *shaking* and *position-checking* actions can be detected visually, *inhaling* and *exhaling* actions do not have any visual cues and require other modalities such as audio. Therefore, we do not include these two actions in our experiments.

Results Our first set of experiments for this dataset with the Snippet Based Trajectory Statistics Histograms is for the *shaking action*. Before we make any comparisons, we need to decide what we consider as a correct detection. We introduce the *overlap ratio* threshold t, which is the ratio of overlap in terms of frames between the detected action and the ground truth. If the detection and the ground truth has a greater overlap than t, then we consider it as a match.

As we see in Figure 4, the performance remains the same until $t = 0.5$. This is also a reasonable threshold value for evaluation. Therefore, we set the overlap ratio threshold $t = 0.5$, and conduct experiments to compare Snippet Based Trajectory Statistics Histograms with other trajectory based descriptors.

Table 1 shows the comparison of trajectory shape, HOG, HOF and MBH features from [27], with Snippet Based Trajectory Statistics Histograms descriptors. Our method

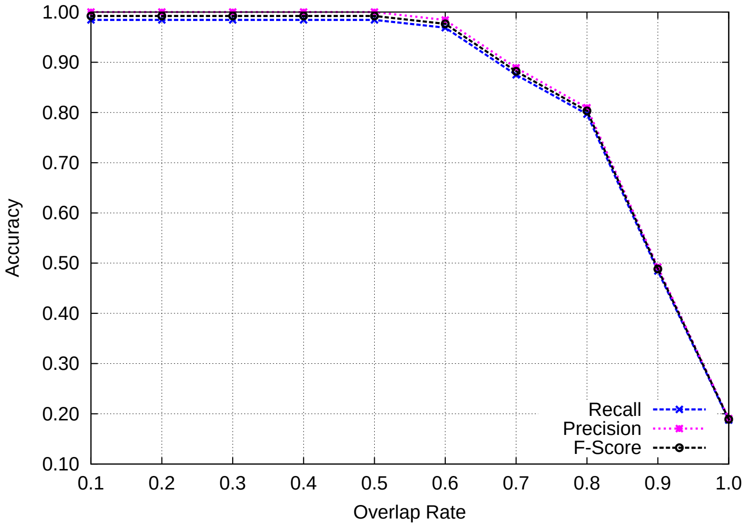


Fig. 4. The effect of overlap rate threshold. Based on this figure, we choose $t = 0.5$ as the overlap rate in our experiments.

Table 1. Comparison of Snippet Based Trajectory Statistics Histograms with other trajectory based descriptors for the *shaking* action of Inhaler Dataset. Trajectory scores are the same as those reported in [29]

	Trajectory	HOG	HOF	MBH	Snippet Hist
Recall	95.31	50.00	**100.00**	87.50	98.44
Precision	91.04	22.70	91.43	71.79	**100.00**
F-score	93.13	31.22	95.52	78.87	**99.21**

has the highest overall F-score, having a perfect 100% precision score. The second-best performing descriptor is HOF, which is not a surprise, because it also mostly encodes the flow information and not the appearance information. If we analyze the shaking action, such as the examples shown in Figure 5, we see that it produces fast and long trajectories. Therefore, it is expected that these descriptors perform very well for this task.

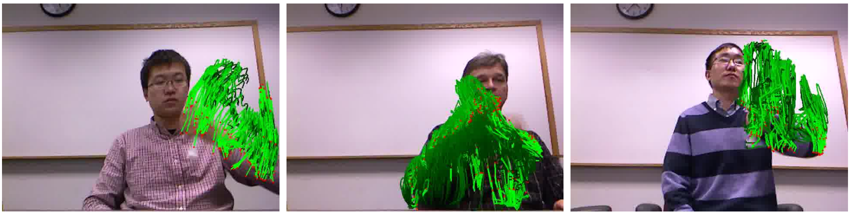


Fig. 5. Regardless of whichever direction user shakes the inhaler device, we get many long trajectories as the result.

Our second set of experiments is done for *position checking*. Note that we have no ways of checking the correct distance between the inhaler and the mouth simply by using the RGB camera, we just try to see if we can detect this action by the user's motion.

Table 2. Comparison of position checking action results. Although RGB-based descriptors are nowhere close to the depth video performance in [29], Snippet Based Trajectory Statistics Histograms still perform better than other trajectory-based descriptors.

	Trajectory	HOG	HOF	MBH	Snippet Hist	Depth
Recall	8.82	8.82	13.24	10.29	30.88	**90.30**
Precision	17.07	12.77	16.36	24.24	18.46	**78.80**
F-score	11.63	10.43	14.63	14.45	23.11	**84.20**

As shown in Table 2, position checking performance is nowhere close to that of [29], which uses the face detector and depth information. However, our goal of this experiment was not to try to improve their result, but rather to compare Snippet Based Trajectory Statistics Histograms with other trajectory-based features. We can see that the Snippet Based Trajectory Statistics Histograms still has the best F-score when compared with the other descriptors, though it is really low.

One of the reasons for such a low score for trajectory-based features in this task is that the motion of moving the inhaler device towards the mouth differs significantly from user to user. Some users move it suddenly, resulting in long trajectories, whereas others move it very subtly, as shown in Figure 6. Therefore, we must make use of depth, face and other information in this task, as done in [29].

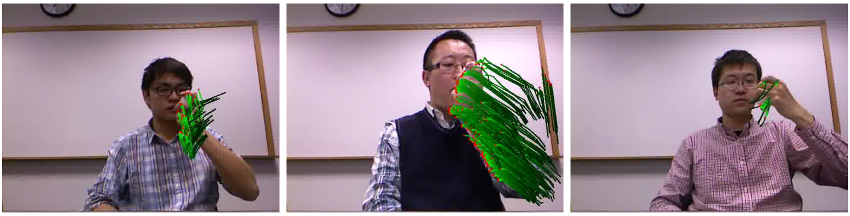


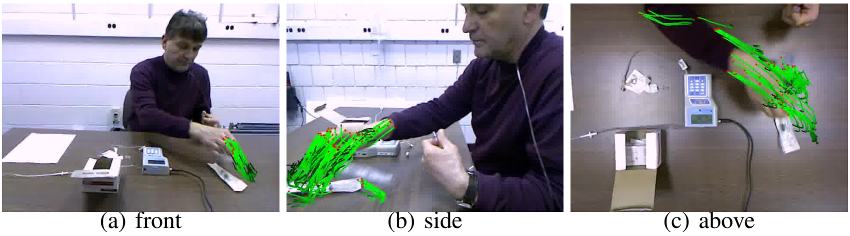
Fig. 6. Position checking is a more subtle action than shaking. The user may slowly or suddenly move the device towards his or her face. As a result, we get trajectories of varying lengths for this action.

4.4 Infusion Pump

Dataset Pump Dataset [2] consists of different users using Abbot Laboratories Infusion Pump device. Each user is asked to perform the operation multiple times, and they may or may not follow the correct order of operation. The dataset is annotated with 7 action classes. There are three different cameras recording each user synchronously from the side, above and front of the user. The dataset has both RGB and depth videos of these recordings, however we only make use of the RGB videos in this paper.

4.5 Results

In order to show the effect of each camera, we report scores separately for each camera in Table 3, as well as taking the best result among 3 cameras and reporting it under the *fusion* column. One of the points that we can observe from these results is that the camera from *above* seems to work better for trajectory-based features than the others. This can be explained by the fact that there are more movements when looked from above compared to side or front during the infusion pump usage (see Figure 7). Since these descriptors depend on trajectories of motion, they work better when there is more motion on camera. Compared to other descriptors, Snippet Based Trajectory Statistics Histograms slightly have better performance in this task.



(a) front (b) side (c) above

Fig. 7. Example of extracted trajectories for the same frame. We can see more motion information from the cameras positioned above and side, compared to the camera positioned in front of the user

Table 3. Results for Pump Dataset. The accuracy performance for each camera is reported separately, and the best performance among three cameras is reported under the *fusion* column.

Actions	Trajectory				HOG			
	front	side	above	fusion	front	side	above	fusion
Turn the pump on/off	65.40	65.40	67.30	67.30	67.13	58.82	72.32	72.32
Press buttons	56.23	63.32	71.11	71.11	56.92	66.78	74.74	74.74
Uncap tube end/arm port	51.04	67.13	63.15	67.13	52.25	70.07	65.74	**70.07**
Cap tube end/arm port	51.04	57.61	66.44	66.44	55.88	68.51	59.34	68.51
Clean tube end/arm port	53.29	58.65	56.57	58.65	56.40	63.32	63.32	63.32
Flush using syringe	47.58	64.53	60.90	64.53	57.27	61.59	77.68	77.68
Connect/disconnect	48.79	56.23	62.63	62.63	54.33	67.47	64.71	67.47
Average	53.34	61.84	64.01	65.40	57.17	65.22	68.26	70.59

Actions	HOF				MBH			
	front	side	above	fusion	front	side	above	fusion
Turn the pump on/off	67.99	65.74	76.99	**76.99**	69.20	68.17	74.91	74.91
Press buttons	45.67	61.59	74.22	74.22	57.44	66.26	72.84	72.84
Uncap tube end/arm port	42.56	59.69	65.57	65.57	57.09	70.07	69.20	**70.07**
Cap tube end/arm port	52.08	72.32	59.52	72.32	52.42	72.84	69.20	**72.84**
Clean tube end/arm port	52.25	62.63	67.13	**67.13**	55.71	56.23	66.78	66.78
Flush using syringe	49.13	73.88	70.59	73.88	54.67	71.80	68.69	71.80
Connect/disconnect	57.27	53.11	60.73	60.73	52.94	71.11	70.59	71.11
Average	52.42	64.14	67.82	70.12	57.06	68.09	70.32	71.48

Actions	Snippet Based Trajectory Statistics Histograms			
	front	side	above	fusion
Turn the pump on/off	75.26	69.03	75.09	75.26
Press buttons	68.34	70.93	76.12	**76.12**
Uncap tube end/arm port	67.13	64.19	67.30	67.30
Cap tube end/arm port	54.33	70.59	66.44	70.59
Clean tube end/arm port	55.02	66.78	65.40	66.78
Flush using syringe	64.01	73.18	82.01	**82.01**
Connect/disconnect	54.33	72.66	72.15	**72.66**
Average	62.63	69.62	72.07	**72.96**

As our last experiment on Infusion Pump Dataset, we train a temporal model based on the order of actions in the training set. It has been shown that this approach improves performance for daily-action recognition applications [7, 11]. We follow the method introduced in [11], and simply train a 2nd-order Markov Chain using the order of operations in order to store the temporal sequence information of the actions.

Table 4. Comparison of temporal sequence model used with different descriptors. We also compare our results with the *ROI-BoW* method introduced in [2]. For this experiment, we only report the best result obtained among 3 cameras for each method

Actions	Trajectory	HOG	HOF	MBH	Snippet Hist	ROI-BoW
Turn the pump on/off	91.52	91.52	90.83	92.39	**97.23**	89.40
Press buttons	79.93	80.28	80.10	79.76	83.91	**88.33**
Uncap tube end/arm port	84.26	85.64	83.56	85.47	**91.35**	65.41
Cap tube end/arm port	84.26	83.91	83.91	84.26	**89.45**	44.55
Clean tube end/arm port	70.24	73.18	77.51	74.05	75.78	**92.02**
Flush using syringe	88.75	88.24	88.06	87.20	92.56	**94.80**
Connect/disconnect	90.14	90.31	88.24	90.14	**92.73**	53.35
Average	84.16	84.73	84.60	84.75	**89.00**	75.41

Based on Table 4, we see that the temporal sequence model has improved the results dramatically. It performs much better than *ROI-BoW* introduced in [2] when used with Snippet Based Trajectory Statistics Histograms. This shows us that the temporal sequence information can be used successfully for medical device usage as well.

5 Conclusion

In this paper we apply Snippet Based Trajectory Statistics Histograms to recognize actions in medical device usage. Our experiments show that they can be used successfully in this domain. This is especially evident for actions with fast motions, such as the *shaking* action. In addition to having a good performance, they do not require any heavy computation and can be easily used for real-time applications. For Infusion Pump Dataset, we also show that storing the sequence information of actions improves the results dramatically when used with Snippet Based Trajectory Statistics Histograms. These results encourage us to believe that the use of medical devices at home can be regulated with the help of technology, saving patients from injuries or fatal incidents.

6 Acknowledgement

This material is based in part upon work supported by the National Science Foundation under Grant No. IIS-1251187. This paper is also partially supported by TUBITAK grant number 112E174.

References

1. Aggarwal, J., Ryoo, M.S.: Human activity analysis: A review. ACM Computing Surveys (CSUR) 43(3), 16 (2011)
2. Cai, Y., Yang, Y., Hauptmann, A.G., Wactlar, H.D.: A cognitive assistive system for monitoring the use of home medical devices. In: Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare. pp. 59–66. ACM (2013)
3. Cardinaux, F., Bhowmik, D., Abhayaratne, C., Hawley, M.S.: Video based technology for ambient assisted living: A review of the literature. Journal of Ambient Intelligence and Smart Environments 3(3), 253–269 (2011)
4. Crispim-Junior, C.F., Bremond, F., Joumier, V., et al.: A multi-sensor approach for activity recognition in older patients. In: The Second International Conference on Ambient Computing, Applications, Services and Technologies-AMBIENT 2012 (2012)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)
6. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: Proceedings of the 9th European conference on Computer Vision - Volume Part II. ECCV'06 (2006)
7. Fathi, A., Farhadi, A., Rehg, J.M.: Understanding egocentric activities. In: Computer Vision (ICCV), 2011 IEEE International Conference on. pp. 407–414. IEEE (2011)
8. FDA: White paper: Infusion pump improvement initiative. http://www.fda.gov/medicaldevices/productsandmedicalprocedures/GeneralHospitalDevicesandSupplies/InfusionPumps/ucm205424.htm (2010)
9. Fleck, S., Straßer, W.: Smart camera based monitoring system and its application to assisted living. Proceedings of the IEEE 96(10), 1698–1714 (2008)
10. Iscen, A., Armagan, A., Duygulu, P.: What is usual in unusual videos? trajectory snippet histograms for discovering unusualness. In: CVPR 2014 2nd Workshop on Web-scale Vision and Social Media (VSM)
11. Iscen, A., Duygulu, P.: Knives are picked before slices are cut: recognition through activity sequence analysis. In: Proceedings of the 5th international workshop on Multimedia for cooking & eating activities. pp. 3–8. ACM (2013)
12. Kitani, K., Ziebart, B.D., Bagnell, J.A.D., Hebert, M.: Activity forecasting. In: ECCV (2012)
13. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. pp. 1–8. IEEE (2008)
14. Lu, Z., Grauman, K.: Story-driven summarization for egocentric video. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 2714–2721. IEEE (2013)
15. Mubashir, M., Shao, L., Seed, L.: A survey on fall detection: Principles and approaches. Neurocomputing 100, 144–152 (2013)
16. Peng, E., Peursum, P., Li, L., Venkatesh, S.: A smartphone-based obstacle sensor for the visually impaired. In: Ubiquitous Intelligence and Computing, pp. 590–604. Springer (2010)
17. Pirsiavash, H., Ramanan, D.: Detecting activities of daily living in first-person camera views. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 2847–2854. IEEE (2012)
18. Rashidi, P., Mihailidis, A.: A survey on ambient-assisted living tools for older adults. IEEE journal of biomedical and health informatics 17(3), 579–590 (2013)
19. Rodriguez, M., Javed, A., Shah, M.: Action mach: a spatio-temporal maximum average correlation height filter for action recognition. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. pp. 1–8. IEEE (2008)

20. Rodriguez, M., Sivic, J., Laptev, I.: Analysis of crowded scenes in video. Intelligent Video Surveillance Systems pp. 251–272
21. Rohrbach, M., Amin, S., Andriluka, M., Schiele, B.: A database for fine grained activity detection of cooking activities. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 1194–1201. IEEE (2012)
22. Roshtkhari, M.J., Levine, M.D.: Online dominant and anomalous behavior detection in videos. In: CVPR (2013)
23. Spriggs, E.H., De La Torre, F., Hebert, M.: Temporal segmentation and activity classification from first-person sensing. In: Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference On. pp. 17–24. IEEE (2009)
24. Sun, X., Yao, H., Ji, R., Liu, X., Xu, P.: Unsupervised fast anomaly detection in crowds. In: ACM MM (2011)
25. Tapu, R., Mocanu, B., Bursuc, A., Zaharia, T.: A smartphone-based obstacle detection and classification system for assisting visually impaired people. In: Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on. pp. 444–451. IEEE (2013)
26. Tenorth, M., Bandouch, J., Beetz, M.: The tum kitchen data set of everyday manipulation activities for motion tracking and action recognition. In: Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on. pp. 1089–1096. IEEE (2009)
27. Wang, H., Kläser, A., Schmid, C., Liu, C.L.: Dense trajectories and motion boundary descriptors for action recognition. IJCV (2013)
28. Wang, Q., Shin, W., Liu, X., Zeng, Z., Oh, C., AlShebli, B.K., Caccamo, M., Gunter, C.A., Gunter, E.L., Hou, J.C., et al.: I-living: An open system architecture for assisted living. In: SMC. pp. 4268–4275 (2006)
29. Wang, Y., Hauptmann, A.: An assistive system for monitoring asthma inhaler usage. In: CMU-LTI-14-002 Technical Report (2014)
30. Winlock, T., Christiansen, E., Belongie, S.: Toward real-time grocery detection for the visually impaired. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. pp. 49–56. IEEE (2010)
31. Wood, A., Stankovic, J.A., Virone, G., Selavo, L., He, Z., Cao, Q., Doan, T., Wu, Y., Fang, L., Stoleru, R.: Context-aware wireless sensor networks for assisted living and residential monitoring. Network, IEEE 22(4), 26–33 (2008)
32. Yoshida, T., Kitani, K.M., Koike, H., Belongie, S., Schlei, K.: Edgesonic: image feature sonification for the visually impaired. In: Proceedings of the 2nd Augmented Human International Conference. p. 11. ACM (2011)
33. Zhao, B., Fei-Fei, L., Xing, E.P.: Online detection of unusual events in videos via dynamic sparse coding. In: CVPR (2011)
34. Ziebart, B.D., Ratliff, N., Gallagher, G., Mertz, C., Peterson, K., Bagnell, J.A., Hebert, M., Dey, A.K., Srinivasa, S.: Planning-based prediction for pedestrians. In: IROS (2009)