

Automated Detection of Unusual Events on Stairs

Jasper Snoek Jesse Hoey Liam Stewart Richard S. Zemel
Department of Computer Science, University of Toronto
10 King's College Rd., Toronto, Ontario, Canada
{jasper,jhoey,liam,zemel}@cs.toronto.edu

Abstract

This paper presents a method for automatically detecting and recognising unusual events on stairs from video data. The motivation is to provide a tool for biomedical researchers to rapidly find and analyse the events of interest within large quantities of video data. Our system identifies potential sequences containing anomalies, and reduces the amount of data that needs to be searched by a human. We apply adaptive background subtraction to segment the person using the stairs, followed by affine flow computation over the segmented region. A hidden Markov model (HMM) is then used to analyse the temporal progression of the affine features. A single HMM is trained on sequences of normal stair use, and a threshold is used to detect unusual events in new data. We also introduce a temporal segmentation method using a conditional random field (CRF). We demonstrate our system on a data set with three persons.

1 Introduction

Stairs have long been the subject of study for architects and designers [19], who attempt to build more ergonomic and safe stairs for different public and private situations. Increasingly, stairs have become a subject of interest for biomedical researchers, who realise that, even with perfect design, stairs are inherently difficult for humans to navigate, and their use will always lead to accidents. The elderly are particularly susceptible to accidents on stairs, as a result of reduced mobility and increased negative impact. This is of special concern to the growing population of elders who wish to age in their homes. Thus, biomedical researchers study the ways in which adverse events happen on stairs, and to identify and possibly predict the causes of these events.

One of the biggest hurdles involved in such research is the gathering of real stair data. Aside from the ethical difficulties of recording stair usage in public or private spaces, there is a technical difficulty imposed by the rarity of ad-



Figure 1. Stairs and overhead camera

verse events. It is estimated that in public staircases, a slip, stumble, trip, or other loss of balance not resulting in a fall occurs once in 2, 222 stair uses, while minor accidents such as falls occur only once in 63, 000 stair uses [1]. The labour intensive process of manually identifying unusual events in stair video data can be avoided with an automated system as we propose herein.

We assume that we have access to a database of stair events on a particular set of stairs, where each stair event consists of a single person entering the stairwell and descending the stairs. Stair events are of two types, *normal* and *anomalous*. In a *normal* stair event, the person descends the stairs with no problems, correctly placing their feet on steps without any loss of balance. An *anomalous* event is one in which the person misses a step at some point in the stair event. A person can miss a step either by completely overstepping, or by catching their heel on the nosing of a step and slipping off onto the next lower step (a *slip*). These are some of the most common small anomalous events on stairs [19]. We assume that our database will consist largely of *normal* stair events, with a small number of *anomalous* events, and we wish to train a system on this data that can classify a video of a new stair event as either a normal event or an anomalous event. The primary goal of our system is to filter a large database, removing a large fraction of stair events which are sure not to contain anomalies. The remaining data could be forwarded to a more complex processing

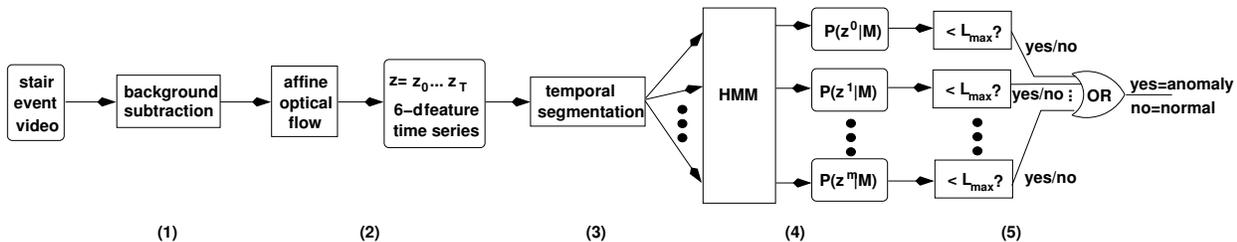


Figure 2. Overview of system. A person’s silhouette in a video of a stair event is extracted using adaptive background subtraction. Affine optical flow is computed over this region resulting in a 6D feature vector for each frame. A temporal segmentation of the stair event results in a set of sequences, each of which are analysed by computing the likelihood of the sequence given a trained hidden Markov model (HMM) and comparing to a threshold L_{max} . Only if all sequences have likelihoods below threshold is the event declared normal, otherwise, it is flagged as anomalous.

stage, to further reduce the number of sequences, and finally to a human for final analysis. Therefore, while our system must not miss any anomalous events, we can afford a reasonable amount of false positive anomalous events. We assume that only a single person is descending the stairs at a time, a limitation that could be overcome with multiple target tracking. We only look at descents, as these are higher risk for adverse events, and are of most interest to biomedical researchers and stairwell designers.

Our system operates in five stages as shown in Figure 2. First, the region of interest (containing the person descending the stairs) is located in each frame using an adaptive background subtraction method. Second, optical flow [3] is computed over the foreground region, and projected to the affine basis. The resulting feature vector forms a time series over each stair descent. Third, the stair event is temporally segmented into sequences, each sequence corresponding to the person descending a single step. In this paper, we compare a manual segmentation with an automatic segmentation from a conditional random field (CRF) [13] model of the stair events. Fourth, we analyse the temporal time series over each sequence using a hidden Markov model [17]. Fifth, we classify each sequence by comparing to a learned threshold, L_{max} , giving a classification (as normal or anomalous) for each sequence. A possible final step, as shown in Figure 2, is to label events as anomalous if any contained sequence is anomalous.

We test our system on a database of events collected from three persons descending an experimental set of stairs in our laboratory, shown in Figure 1. For each test subject, we gathered normal stair events and events where the person either slipped or overstepped. We trained and tested our system on different combinations of subjects’ data, and looked either at the classifications of sequences, or of entire events. As described in Section 4, our current results, averaged over subjects, are a 1% miss rate and a 22% false

positive rate for a system trained and tested on a single person, and a 6% miss rate and a 35% false positive rate on a system trained on two people and tested on a third.

2 Previous Work

There is much work on detecting anomalous behavior in video in the context of visual surveillance [7] or user modeling [5, 14]. However, these approaches use coarse features such as positions and velocities of persons within a scene and attempt to characterise trajectories. A larger body of computer vision research has looked into modeling the motion of the human body in fine detail. Periodic motion of walking figures is analysed in [6] by computing self-similarity of a segmented image region with itself over multiple time scales. The Fourier transform of the resulting correlations gives indications of the periodicity of the motion. Human gaits were used for person identification in [15] by analysing the spatial moments of the optical flow across an image where a person is moving. The patterns of the relative phases of these moments were then compared across multiple individuals. Similar features could also be applied in the work we describe here. The motion history (MHI) [4] is a descriptor of temporally localised image changes. However, these works do not attempt to recognise anomalous events and do not look at motion on stairs.

Our work uses mid-level optical flow features, analysing the motion of a single person’s body performing a task, and looks at anomalous events on stairs in particular. Relatively little work has been done on characterising human motion on stairs. Notable exceptions are work done on motion capture data of persons ascending and descending stairs, in which recovered joint angles are mapped to a subspace which can be used for synthesis. However, this work does not use video and does not attempt recognition of unusual events. Human gaits on staircases were analysed in [2] by

fitting a skeletal model to the view-based human form, and then modeling the joint angles as a dynamical system. This was used to classify gaits such as walking, running, and descending stairs, but no work was done on recognising unusual events within each of these motion types. An interesting study in [11] used a camera mounted above a side-by-side public stairway and escalator to implement a prompting device that would encourage people to use the stairs if they were about to use the escalator. Background subtraction was used to determine if people were using the escalator or the stairs, and the work also addressed some issues of multiple persons on the stairs. However, there is no recognition of unusual events.

3 Stair Event Classification

3.1 Background Subtraction

We use a simple adaptive background subtraction technique where we threshold the absolute difference between a new image at time t , $I_t(x, y)$, and a 'reference image', $A(x, y)$, containing only the background. This technique is very sensitive to changing background conditions, and so the reference image is updated after each frame by taking a weighted average of all previous images in a sequence, with a learning rate of α_b :

$$A_t(x, y) = (1 - \alpha_b) * A_{t-1}(x, y) + \alpha_b * I_t(x, y).$$

In our experiments, we used $\alpha_b = 0.8$ for the first 100 frames, then 0.0005 afterwards.

This technique suffers from a number of factors, the most significant of which are shadows and specularities. A number of methods exist which attempt to deal with these issues. These include using difference in depth from stereo information to segment the background [12], multi-component systems [20], and formulating probabilistic models of background pixels which are then segmented using mixtures of Gaussians [9, 21]. We opted instead, due to its simplicity and good results, for a method where we perform a second round of background subtraction on the result of the initial background subtraction, but in the hue channel of the HSV color space. We found that this removed the majority of shadow pixels because the hue is less sensitive to changes in brightness than the RGB color space. Finally, we removed remaining noise by convolving with a gaussian kernel and finding the largest connected component. Figure 3 shows examples of the background subtraction.

3.2 Optical Flow Features

We use optical flow as our primary identification feature as it is independent of the overall lighting and color

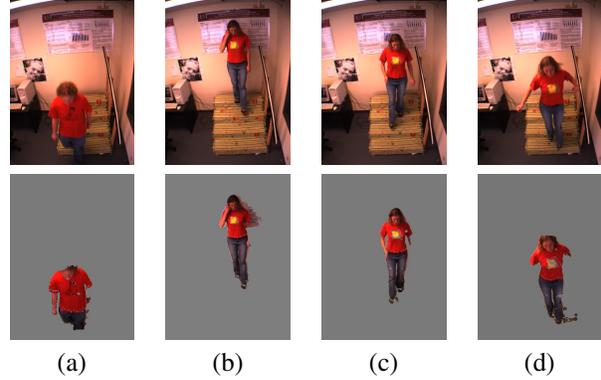


Figure 3. Four examples of background subtraction (a) Subject 2 (b)-(d) Subject 3

conditions. Optical flow is the motion $\mathbf{v} = \{v_x, v_y\}$ at a pixel between subsequent frames of video and is computed by finding solutions to the brightness constancy constraint, $f_t = \nabla \mathbf{f} \cdot \mathbf{v} = f_x v_x + f_y v_y$, where f_t, f_x, f_y are the temporal, horizontal spatial and vertical spatial image derivatives. We use the robust dense optical flow method of [3], which finds locally consistent solutions to the brightness constancy equation using a robust error norm. This method is fairly insensitive to outliers (e.g. due to violations of brightness constancy), and produces smooth flow fields due to the regularising effect of enforcing local consistency.

To reduce the dimensionality of the optical flow, we look at the overall distribution of flow vectors at varying levels of spatial frequency across a region of interest defined by the background subtraction. In this paper, we use only the affine component of the flow field, which is a description of the flow as a planar spatial function with six parameters, $\mathbf{z} = \{a, b, c, d, e, f\}$ as follows:

$$v_x = ax + by + c \quad v_y = dx + ey + f$$

In fact, for the results we present in Section 4, we only use the lowest order (means): c and f . It was found that, for the level of detail we were pursuing, this simplest representation was sufficient.

3.3 Temporal Segmentation

The full stair events are temporally segmented into step sequences, such that each sequence contains the motion of one foot from the time it begins to leave one step until it lands firmly on another. The temporal segmentation is done separately from the HMM-based event recognition, and is either manual (by one of the authors) or automatic using a conditional random field (CRF) [13], as we now describe.

The training data for the CRF method consists of the set of labels for each frame, $Y = \{y_1, \dots, y_T | y_i \in \mathcal{Y}\}$, and the

observations, $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_T\}$. In our case, the labels are binary and correspond to which foot is currently in motion, so $\mathcal{Y} = \{left, right\}$. A CRF models the joint distribution of the *labels* conditioned on the observations as a log-linear combination of feature functions. The most common model for sequential data is a linear chain model, in which the feature functions relate adjacent pairs of labels and the observations, and are homogeneous (weights are shared across the sequence). We define our features to be

$$f_{i,j}(y_t, y_{t-1} | \mathbf{Z}) = \delta(y_t = i, y_{t-1} = j) \theta_{i,j} \cdot \mathbf{z}_t$$

for all $i, j \in \{left, right\}$, where $\theta_{i,j}$ is a learned weighting coefficient. In addition, we have a set of similar features $f_i(y_t | \mathbf{Z})$ that are used for the first element of a sequence. The joint probability of a sequence is thus

$$P(Y | \mathbf{Z}) = \Omega \exp \left\{ \sum_i f_i(y_1 | \mathbf{Z}) \sum_{t=2}^T \sum_{i,j} f_{i,j}(y_t, y_{t-1} | \mathbf{Z}) \right\}$$

where Ω is the normalization constant, which, along with the singleton and pairwise marginals, can be computed efficiently using belief propagation. We use a limited memory quasi-Newton algorithm to train the CRFs in a penalized (weight decay) maximum likelihood framework.

This CRF model is trained on stair events with manual temporal segmentations. We augment each observation with all observations in a window of five around it (chosen using 3-fold cross-validation) and normalize the observations to have zero mean and unit standard deviation. This prevents abnormally large observation values from dominating the features and negatively impacting training. At test time, the transformation from the training data is applied to the testing data and Viterbi decoding is used to obtain a labeling.

3.4 Event Classification

Once we have extracted our low dimensional (affine) features from the optical flow over the foreground region in each frame and have temporally segmented a stair event into a number of step sequences, we wish to classify the step sequences as being either normal or anomalous. To do so, we use a hidden Markov model, or HMM. Hidden Markov models have enjoyed great popularity in computer vision based event recognition due to their flexibility and generality (see e.g. [18, 8]). This computer vision work grew from successes from using HMMs in speech recognition [17]. A hidden Markov model is a probabilistic temporal model $\{\mathcal{S}, \mathbf{Z}, T, B\}$, where \mathcal{S} is a finite set of states, \mathbf{Z} is a continuous observation space, $T : \mathcal{S} \rightarrow \mathcal{S}$ is a transition function giving the probability of transitioning from state s at time t to state s' at time $t + 1$, $T(s, s') = Pr(s' | s)$, and $B : \mathcal{S} \rightarrow \mathbf{Z}$ is an observation function giving the probability of observing observation feature vector \mathbf{z} given state s :

$B(s, \mathbf{z}) = Pr(\mathbf{z} | s)$. The observation function for a continuous space is parameterised using a Gaussian:

$$B(s, \mathbf{z}) = \mathcal{N}(\mathbf{z}; \mu, \Sigma),$$

where μ, Σ are the mean and covariance matrix of the Gaussian component model.

We train the HMM using the expectation maximization algorithm, as implemented in the BNT toolbox [16]. We used 8 hidden states and diagonal covariance matrices, and initialised the EM algorithm randomly. We use the standard forward algorithm to evaluate the likelihood of a new sequence given a trained HMM. The number of hidden states was chosen by evaluating performance of the method using different settings (see Section 4).

For each experiment, we split the data into training and test sets, as detailed in Section 4. We use the manual temporal segmentations for the training set and either the manual or automatic segmentations in the test set. The CRF segmenter was trained using the first 12 components of the Zernike polynomial basis (an extension of the affine basis, described in [10]). We then perform a cross-validation procedure on the training data. We remove all anomalous sequences and a single normal sequence, s_i , from the training set and train the HMM, M , on the remaining normal sequences. We then compute a cost for the validation set for each value of L

$$C_i(L) = C_f^v g(\log P(s_i | M_v), L) + \frac{C_m^v}{N_a} \sum_{j=1}^{N_a} (1 - g(\log P(s_j | M_v), L)) \quad (1)$$

where N_a are the number of anomalous sequences in the training (and hence validation) set, s_j is one of these anomalous sequences, M_v is the trained HMM and $g(p, t) = 1$ if $p < t$ and 0 otherwise. The constants C_f^v, C_m^v give the relative cost of a false positive anomaly (a normal sequence classified as an anomaly) and a missed anomalous sequence, respectively. We choose $C_m^v = 3C_f^v$: a miss is three times worse than a false positive, as prescribed by our aim of filtering normal steps out of the data set, while preserving anomalous steps. We then choose L_{max} to minimise the summed cost over all validation sets

$$L_{max} = \arg \min_L \sum_{i=1}^{N_r} C_i(L)$$

where N_r is the number of normal training sequences.

Finally, we train an HMM, M_t , on all the normal sequences in the training data. We evaluate the test sets for the manual segmentations by counting the fraction of false positives, $N_n^{-1} \sum_{k=1}^{N_n} g(\log P(s_k | M_t), L_{max})$, where s_k is the k^{th} normal test sequence and N_n is the number of normal test sequences, and the fraction of misses

$N_a^{-1} \sum_{l=1}^{N_a} (1 - g(\log P(s_l|M_t), L_{max}))$, where s_l is the l^{th} anomalous test sequence and N_a is the number of anomalous test sequences. In the WEAK generalisation experiments, these results are averaged over all validation sets.

We compared our HMM method to a simple polynomial ridge regression method, in which order- K polynomials were fit to normal and abnormal step sequences in the training set and used to classify the test data. In the simplest, $K = 0$ case, this method classifies the sequences based only on their mean values, and is akin to a simple threshold in the temporally averaged mean flow. Higher orders of K give more complex fitting capabilities.

4 Experiments and Results

This section describes the results we obtained on a small dataset of three subjects descending an experimental set of four stairs in our laboratory, as shown in Figure 1. A Point Grey Research BumblebeeTM camera with a 4mm focal length was mounted on the ceiling at a perpendicular distance of approximately 280cm from the nosing of the center step in the stairway. The camera is visible in the top right corner of Figure 1. We used the RGB color images at 640×480 resolution from the reference (left) camera¹. The subjects descended the stairs in three sets of events. In the first set, they descended normally. In the second set, they missed one step completely (an *overstep*). In the third set, they slipped off one step (a *slip*). In total, subject 1 had 165 normal step sequences, and 30 anomalous step sequences, subject 2 had 85 normal and 26 anomalous, and subject 3 had 86 normal and 26 anomalous.

We did four separate types of experiments. The first three experiments tested weak generalisation, in the training set included examples of the person in the test set. These experiments are labeled Ti -WEAK, where $i \in \{1, 2, 3\}$ is the number of subjects in the training set. In the fourth experiment, $T2$ -STRONG, no examples of the test subject are included in the training data, testing strong generalisation across people. It is this fourth experiment that is the most relevant to our final system: we want to be able to flag an anomalous event occurring for a never before seen person descending the stairs. However, it is also the most difficult: the types of normal motion exhibited by the unseen person will not be modeled by the HMMs and will more often be flagged as anomalous. In the WEAK generalisation experiments, we split the data into two parts: a training set comprised of all but one of the normal stair events and only one stair event with an anomaly (an overstep or slip), and a test set with the one left out normal stair event and all but one stair event containing an overstep or slip step sequence. This leave-out procedure is repeated N times, where N is

¹Stereo information was not used in this study, but could play an important role in future work.

the number of normal step sequences in the training data. In the STRONG generalisation experiments, the training set is all the data from 2 subjects, and the test data is all the data from the third subject.

Figures 4–6 show examples of stair events for two different subjects. In each figure, the top row shows original images for key frames with the foreground region marked. The second row shows the corresponding optical flow fields. The bottom rows show the affine coefficients for vertical (top) and horizontal (bottom) flow. The solid lines show the mean flow. Figure 4 shows a normal step. We can see the periodic motion in both the horizontal and vertical flow components. Figure 5 shows an example of an overstep on the 3rd step. In this case, the periodicity is not as clear, but we see a large motion towards the end of the sequence. Figure 6 shows an example of a slip for a different subject on the 3rd step. We see large motions towards the end of the sequence, and less regular periodic structure over the event.

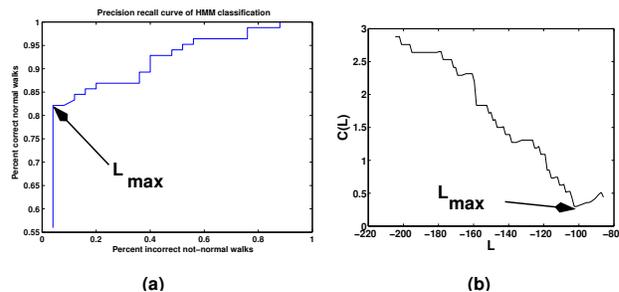


Figure 7. (a) Precision recall curve (b) cost function $C_i(L)$. Plots show the value of L_{max} .

Figure 7(a) shows the precision recall curve (the relative values of the two terms in Equation 1), and Figure 7(b) shows the cost $C_i(L)$ as a function of L for one test subject for one train-test data split, i , for a $T1$ -WEAK experiment. The cost function is used to set the threshold L_{max} . We can see in Figure 7(a) how this impacts the tradeoff between misses and false positives.

We set the number of hidden units in our HMMs by testing different numbers on the $T1$ -WEAK experiments, as shown in Table 1. The results are fairly consistent between 4 – 10, and we use 8 in our experiments.

units	1	2	4	6	10	16
train	.96	1.0	1.0	1.0	1.0	1.0
test	.80	.81	.83	.84	.84	.79

Table 1. Results for differing numbers of hidden units (manual segmentations, $T1$ -WEAK).

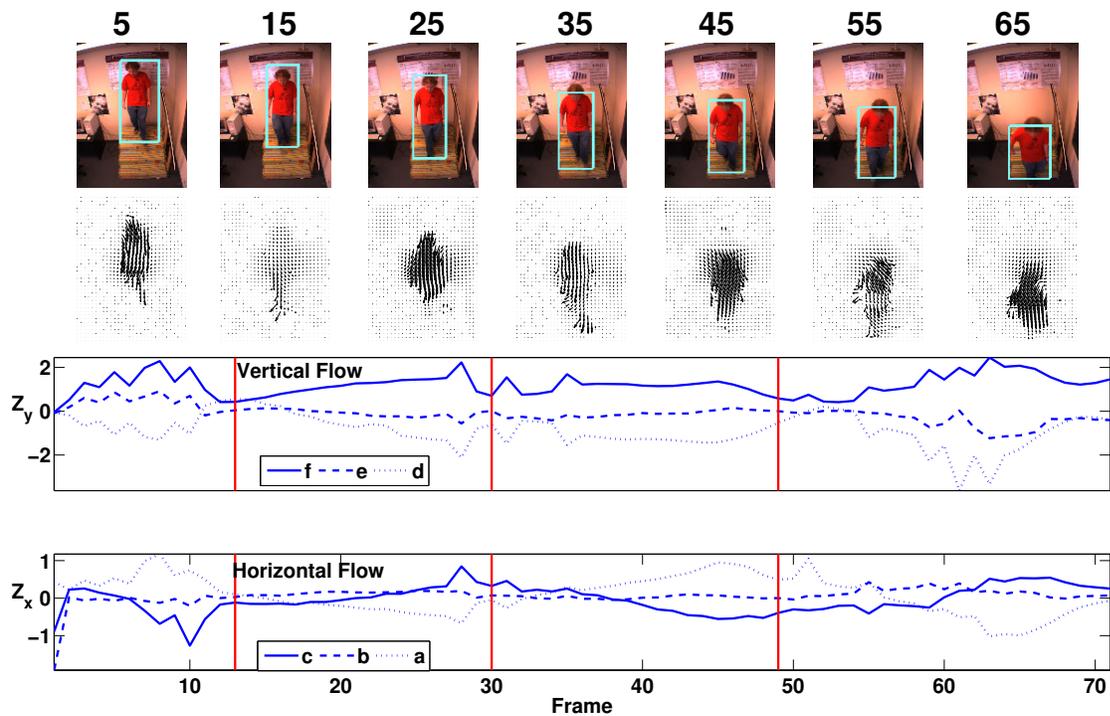


Figure 4. Example of a person descending stairs normally. Key frames and optical flow fields are shown along the top two rows, while the bottom plots show the vertical (above) and horizontal (below) affine flow coefficients. The vertical lines show the manual temporal segmentations.

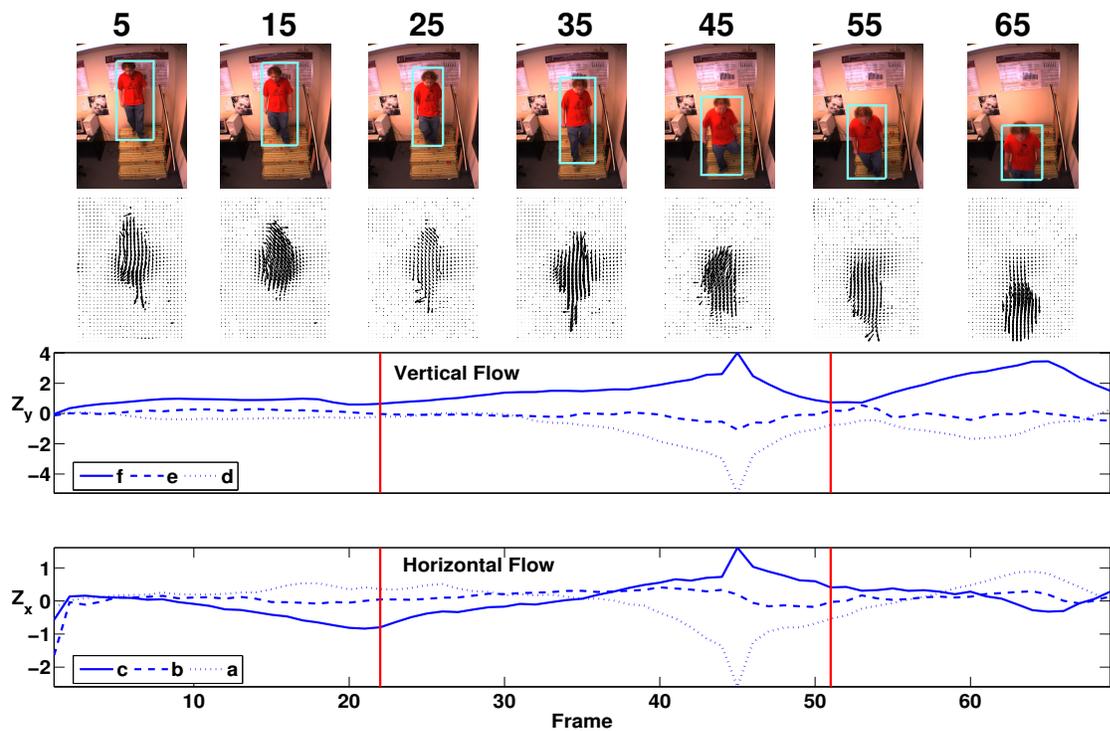


Figure 5. Example sequence of an anomalous event - an overstep in frames 22-51

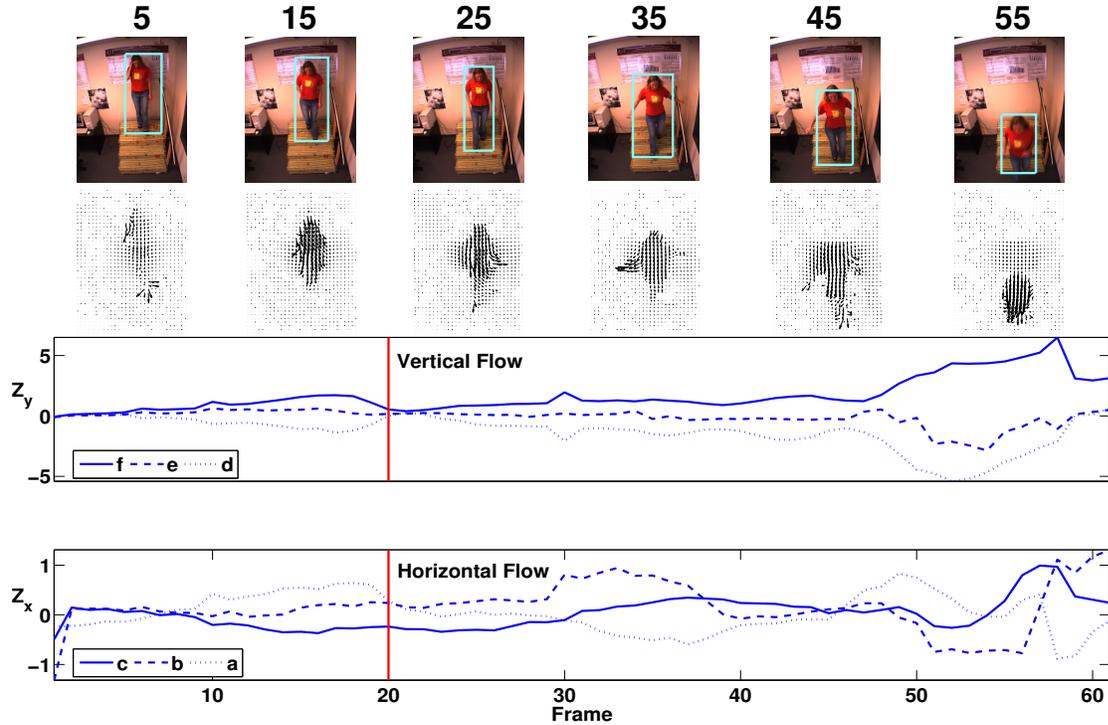


Figure 6. Example sequence of an anomalous event - a slip in frames 20-62

Exp. type	test subj.	missed anom.(%)	False +ve anom.(%)
T1-WEAK	1	0.0	15.9
	2	4.0	31.2
	3	0.0	23.8
	avg	1.2	21.8
T2-WEAK	1,2	5.5	29.6
	1,3	3.6	32.9
	2,3	0.9	41.2
	avg	3.4	33.5
T2-STRONG	1	0.0	43.6
	2	7.7	30.2
	3	11.5	25.8
	avg	6.1	35.6
T3-WEAK	1,2,3	4.9	38.8

Table 2. HMM results (manual segments).

Table 2 shows our results for manual temporal segmentation of the training data, for each of the experiment types and subjects, using only the lowest order (mean) flow. Also shown are the averages across all test sets for each experiment type. We see that, for the T1-WEAK experiments, we can get very low miss rates, and reasonable false positive rates. For comparison, Table 4 shows the the miss

test subj.	segm. error(%) (train)	segm. error(%) (test)	missed anom. (%)	False +ve anom. (%)
1	3.6	31.1	10.6	58.4
2	6.0	17.0	0.0	70.8
3	4.8	32.1	23.3	68.0

Table 3. Results for CRF segmentations of the test data for T2-STRONG experiments

and false positive rates for the polynomial regression for the T1-WEAK experiments. We see the results are worse, yielding unacceptable miss rates at similar false positive rates.

When generalising to more than one person in both training and test sets, T2-WEAK and T3-WEAK, we see similar results, with slightly more false positives and misses. The experiments on unseen subjects (T2-STRONG), are again slightly worse, giving over 10% misses on one test subject, but only 5.5% and 3.6% on the other two. These experiments show that we can reduce the data set size by almost 65% on average across persons, while only missing 6% of the anomalous events. The false positive rates are similar to those seen in the T2-WEAK experiments. The additional misses are due to variability between persons in stair descent styles or speeds, for example. Our methods rely on a

test sub- ject	K=0		K=1		K=4	
	miss anom (%)	F+ve anom (%)	miss anom (%)	F+ve anom (%)	miss anom (%)	F+ve anom (%)
1	33.3	17.6	50.0	26.7	36.7	29.1
2	80.7	50.0	42.3	31.4	57.7	27.9
3	34.6	40.0	34.6	36.6	42.3	29.4

Table 4. Polynomial regression results for T1-WEAK (manual segmentations).

good estimate of the between-person variance, and so would improve with data from more people. Our *a-priori* methods for setting the costs C_f and C_m also need further exploration. Another avenue for research are more complex representations of motion, such as skeletal models. The filtering step we present here, however, would be appropriate as a front-end for such more intensive analyses.

Table 3 show the results when using CRF segmentations of the test data. The results in this case are evaluated on a per-frame basis, and thus are showing, as misses, the total number of frames in an anomalous sequence that were labeled as normals, and, and false positives, the total number of frames in normal sequences that were labeled as anomalous. Also shown are the per-frame error rates of the CRF segmentations when compared to the manual ones. In this case, we only ran the T2-STRONG experiments, and found that (as expected) the results were worse. In some cases, we see low miss rates at the cost of high false positive rates. The high false positive rates could be due to many small segments which are found to be classified as anomalous, but actually are part of a longer normal sequences.

5 Conclusions

We have demonstrated a method for filtering a large database of video sequences of people descending stairs for anomalous events. This type of method has the potential to be invaluable to biomedical researchers interested in the causes and types of stair accidents. We used optical flow features and a hidden Markov model to detect unusual events in data, and experimented with a conditional random field for automatic temporal segmentation of the data. Our results demonstrate that our method could be used as an effective first filtering step to remove large quantities of normal steps from data. Our current work is investigating how to further filter the reported anomalous events by more complex visual analyses.

Acknowledgements: We thank Brian Maki, Alex Mihailidis, Jon-Paul Lobos, and Jennifer Boger. This work was supported in part by a grant from the Canadian Institutes of Health Research.

References

- [1] J. Archea, B. Collins, and F. Stahl. *Guidelines for Stair Safety*, volume 120 of *NBS Building Science Series*. National Bureau of Standards, Washington D.C., 1979.
- [2] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. In *Proc. of CVPR 2001*, volume II, pages 52–58, Hawaii, USA, December 2001.
- [3] M. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise smooth flow fields. *CVIU*, 63(1):75–104, January 1996.
- [4] A. F. Bobick and J. W. Davis. The recognition of human movement using temporal templates. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 23(3):257–267, March 2001.
- [5] M. Brand and V. Kettner. Discovery and segmentation of activities in video. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 22(8):844–851, August 2000.
- [6] R. Cutler and L. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 22(8):348–362, 2000.
- [7] J. Fernyhough, A. Cohn, and D. Hogg. Constructing qualitative event models automatically from video input. *Image and Vision Computing*, 18(9):81–103, 2000.
- [8] A. Galata, N. Johnson, and D. Hogg. Learning variable-length Markov models of behavior. *CVIU*, 81(3), 2001.
- [9] W. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in a site. In *Proc. of CVPR*, Santa Barbara, CA, 1998.
- [10] J. Hoey and J. J. Little. Representation and recognition of complex human motion. In *Proc. of CVPR*, pages 752–759, Hilton Head, SC, June 2000.
- [11] J. A. Hyman. Computer vision based people tracking for motivating behavior in public spaces. Master’s thesis, Massachusetts Institute of Technology, 2003.
- [12] Y. Ivanov, A. Bobick, , and J. Li. Fast lighting and independent background subtraction. In *Proc. IEEE Workshop on Visual Surveillance*, 1998.
- [13] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. of ICML-01*, pages 282–289, 2001.
- [14] L. Liao, D. Fox, and H. Kautz. Learning and inferring transportation routines. In *Proc. AAAI*, pages 348–353, 2004.
- [15] J. J. Little and J. Boyd. Recognizing people by their gait: the shape of motion. *Videre*, 1(2), Winter 1998.
- [16] K. Murphy. The Bayes net toolbox for Matlab. In *Proc. of Int. of Computing Sci. and Stats.*, volume 33, 2001.
- [17] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE*, 77(2):257–296, February 1989.
- [18] T. Starner and A. P. Pentland. Visual recognition of american sign language using hidden Markov models. In *International Workshop on Automatic Face and Gesture Recognition*, pages 189–194, Zurich, Switzerland, 1995.
- [19] J. Templer. *The Staircase: Studies of Hazards, Falls, and Safer Design*. MIT Press, January 1995.
- [20] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. of ICCV*, Corfu, 1999.
- [21] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 19(7):780–785, 1997.