GEI + HOG for Action Recognition

Tenika Whytock
tpw3@hw.ac.uk
Alexander Belyaev
http://www.hw.ac.uk/~belyaev
Neil Robertson
http://home.eps.hw.ac.uk/~nmr3

School of Engineering & Physical Sciences
Heriot-Watt University
Edinburgh, UK

Abstract

This paper demonstrates the benefit of applying the Gait-Energy Image (GEI) [15] and Histograms of Oriented Gradients (HOG) [12] descriptors for action recognition. Multi-class Support Vector Machine (SVM) classification show promising results at 100% using leave-one-out cross validation. Furthermore, this technique gains 27° viewpoint tolerance and robustness to occlusions, clothing and carrying condition variations. The contribution of this paper is two-fold. The first employs a traditional gait recognition representation alongside HOG descriptors for action recognition, while the second decomposes actions into static and dynamic classes for superior performance and reduced processing time.

1 Introduction

Action recognition is a hot topic in the computer vision community, where various techniques achieve high success despite the introduction of more complex and realistic datasets. Potential applications include visual surveillance, human-computer interaction and video indexing. Humans have investigated action performance since the 15th Century [18] starting with anatomical studies and progressing to biomechanical research, cinematography and motion perception before emerging to the computer vision techniques employed today. While humans can easily label actions, replicating performance with computer vision is challenging. There are a number of surveys available on action recognition, where recent examples include Poppe [25] and Aggarwal and Ryoo [1].

This paper presents a new action recognition technique combining the Gait-Energy Image (GEI) representation, Histograms of Oriented Gradients (HOG) descriptors and multi-class SVM classification. This technique for action recognition assumes the following: all subjects are healthy, viewed from the side or 0° view, single person performing a single action per sequence and actions are pre-classified as static or dynamic.

While all actions follow the same fundamental pattern of movement, variations relating to magnitude and timing occur, and unlike gait recognition, action recognition generalises over such variations forming an action label. Action labels resembling broad natured verbs

are desired to be comprehensible to humans without action-specific knowledge.

Covariate factors are the challenges faced during action recognition which can affect the appearance and performance [2] of an action and cause decreasing performance. Action recognition relies on discriminative features, and covariate factors tend to determine the best suited representation, features and classification technique. Common covariate factors employed within datasets are clothing, carrying condition, occlusion, lighting, ground surface and viewpoint variations. It is desirable to test an algorithm initially on a dataset containing no covariates, as low performance here suggests the lack of robustness and ability.

Contributions

This paper investigates the performance of the GEI, traditionally employed for gait recognition, alongside HOG descriptors to form a global grid-based approach [25] to action recognition. HOG is traditionally applied to RGB or grey-scale images which contain wide spread gradients, however the GEI produces a space- and time-normalised figure from a binary silhouette sequence where areas containing little to no gradients exist. With this, a number of gradient filters, cell and bin sizes and SVM kernels, including those commonly employed for HOG descriptors, are evaluated. This paper performs action recognition based on the combination of the GEI, HOG and SVM, where the contribution is as follows:

- deployment of the GEI due to the single compact 2D representation over a silhouette sequence, reduced processing time and noising mitigating attributes, combined with HOG descriptors employing various gradient filter, cell and bin sizes
- performance evaluation of all actions versus decomposition into static and dynamic action classes for reduced processing time

The paper is organised as follows: Section 2 presents related work to the applied techniques and state-of-the-art approaches in action recognition, Section 3 describes the action recognition framework, covering GEI representation and HOG descriptors, Section 4 describes the dataset employed, action decomposition, SVM techniques and the results from the best performing HOG parameters and SVM kernels, and Sections 5 and 6 discuss and conclude the results given all variables respectively and highlights details of future work.

2 Related Work

Action recognition is a popular research topic and performance remains high despite introduction of more challenging datasets. The following techniques are highlighted based upon popularity, performance and state-of-the-art.

A similar global representation is the enhanced GEI (EGEI) [21] which is derived from the GEI. The EGEI enhances the dynamic sections due to their discriminative power, where 2D Principle Component Analysis is employed for dimensionality reduction and nearest neighbour classification is employed for action recognition. Despite the attractive performance, further robustness evaluation is required. Fusion of motion and shape features, by Sun et al. [29], is presented for gait recognition as opposed to action recognition. Motion features based on shape variation-based (SVB) frieze pattern provide robustness to carrying condition, while shape features employ the GEI. Dynamic time warping (DTW) is required

to match the motion features due to sequence length variation, and dimensionality reduction is applied using coupled subspaces analysis (CSA) and discriminant analysis with tensor representation (DATER). Despite individual features performing well, fusion is more beneficial for robustness to covariate factors. To combat viewpoint variance, Lin et al. [22] compute three GEIs from different views for each action. A less time intensive minimum incremental coding length (MICL) classification approach is applied to each GEI where a majority wins technique concludes the action class. Furthermore, the size of GEI is analysed to investigate performance given the trade off between reduced space and processing time versus data quality. The direction of GEIs is also questioned where those same facing produce superior results. While this technique has competitive results, further analysis is required on alternative covariate factors. Closely related representations to the GEI are the Motion-Energy Image (MEI) and Motion-History Image (MHI) [6], each producing alternative features shown in Table 1. The MEI and MHI representations demonstrate where and how motion occurs respectively and contain dynamic, and dynamic and temporal features respectively. In comparison, the GEI contains static and dynamic features which are desirable given the temporal aspect may only be matched during specific stages of action performance, and static information alone only shows the overall shape of action.

Features	MEI	MHI	GEI
Static	No	No	Yes
Dynamic	Yes	Yes	Yes
Temporal	No	Yes	No

Table 1: Comparing Motion- and Gait- Energy and History Images

Local descriptors, implemented by Dalal and Triggs [12], show HOG descriptors detecting humans in still images. Later expansion enables classification of actions [19] and gender [7]. Dalal and Triggs [12] claim an accurate gradient filter with no smoothing is essential for person detection due to the associated lower miss rate. The best performing gradient filter for the application is a 1D centred mask with a miss rate of 11%. However of those filters evaluated, the Sobel filter ranks last, with a 3% higher hit rate. HOG descriptor performance may vary based gradient filter, cell and bin size, application and dataset. Laptev et al. [19] use a Hessian detector and HOG and Histograms of Optical Flow (HOF) descriptors alongside a Bag-of-Features (BoF) representation to perform action recognition on simple and realistic datasets. Results indicate HOF outperforms HOG on both simple and realistic datasets, where descriptor combination produces competitive results. Performance is lower on realistic datasets due to complexity, since content includes sequences from television or film sources where body views are commonly partial, whereas simple datasets show full body views. Such local representations are attractive since person detection is not required, however initial application here is geared towards full body views where background modelling is permitted.

3 Action Recognition Framework

The framework for action recognition is composed of action representation and descriptor computation. While global representations produce a quantity of features, sensitivity to noise, occlusion and viewpoint occurs; however deployment of global grid-based representations combats the aforementioned pitfalls and requires a global representation with a local

descriptor. While the GEI is traditionally employed for gait recognition, the associated attributes lend themselves to action recognition. The HOG descriptor is a popular technique not restricted to action recognition, and the gradient filter and cell and bin size require investigation given the alternative representation.

GEI Representation

The GEI, by Han and Bhanu [15], is a global appearance representation successfully employed for gait recognition, and has attributes attractive for action recognition. The GEI, shown in Figure 1, reflects the overall silhouette structure and corresponding changes during action performance and expresses static and dynamic information. The GEI is defined by equation 1:

$$G(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y) \tag{1}$$

where $G(x,y)$ is the GEI, N is the number of frames, t is the frame number, x and y are the 2D spatial image coordinates and $B_t(x,y)$ is the silhouette.

The representation is named such as each silhouette is the space-normalised energy image of the action at a specified time and the time-normalised accumulative energy image of the silhouette during action performance. The higher intensity pixels indicate static areas, while lower intensity pixels highlight dynamic portions of the performed action respectively.

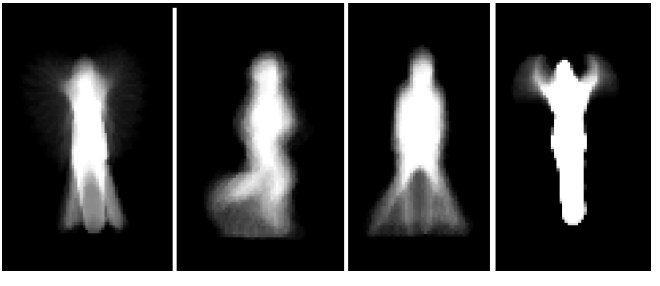


Figure 1: Exemplar Gait-Energy Images created from the Weizmann Action Dataset showing from left to right: jumping jack, run, gallop sideways and two-hand wave.

The GEI has a number of attractive attributes which lend themselves to action recognition. Compared to binary silhouette sequences, the GEI is a compact 2D image reducing spatial and processing requirements. Furthermore, noise mitigation occurs due to time normalisation [15]. Image pre-processing is required for GEI construction and requires background subtraction, size normalisation and horizontal alignment of silhouettes. While robustness to short term occlusion exists, the GEI benefits from construction at a view which expresses the most dynamic information, which for walking is a side view where the swing of the arms and legs is visible. Furthermore, the GEI size has shown to affect performance [22], and may be of interest considering data resolution.

HOG Descriptor

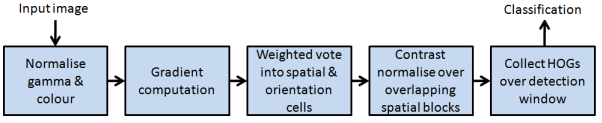


Figure 2: Process for computing HOG descriptors, image adapted from [12].

The HOG descriptor, proposed by Dalal and Triggs [12], enables human detection in still images using the process illustrated in Figure 2. HOG is a local descriptor and is commonly employed for local representations. Primary variables of interest for action recognition are gradient filter and cell and bin size. Various gradient filters are available ranging in accuracy and computational complexity, where performance is linked to gradient computation and large parameter values produce high dimensionality feature vectors.

HOG descriptors are computed in five stages. Gamma/colour normalisation can be performed, but has little impact due to subsequent normalisation. The image is initially divided into cells prior to gradient computation. Each pixel contributes a weighted vote using the $L^2 - norm$ to the gradient orientation, and votes are aggregated into orientation bins over cells. Cells can be rectangular or radial and orientation bins are evenly spaced over 0°-180° or 0°-360°. Orientation voting is based on gradient magnitude values. Gradient magnitude strength varies with illumination and contrast, where contrast normalisation is vital for increased performance and invariance to illumination and shadows. Normalisation groups cells into blocks, where blocks overlap and cells contribute multiple times to the final descriptor, subsequently increasing performance. Two block geometries can be employed, rectangular (R-HOG) where square or rectangular blocks are partitioned into square or rectangular cells, or circular (C-HOG) where circular blocks are partitioned into cells in a log-polar fashion. R-HOG and C-HOG are similar in nature to Scale-Invariant Feature Transform (SIFT) and Shape Contexts respectively.

Most HOG descriptors are applied to RGB or grey-scale images, however the GEI does not contain wide spread gradients. Considering this difference, a range of cell and bin sizes and gradient filters of varying accuracy and computational complexity require investigation. Henceforth, the results are based on HOG [24] for grey-scale images using the following properties: six gradient filters, ten cell and bin sizes and R-HOG.

Estimating Image Derivatives

Dalal and Triggs [12] present HOG performance with deployment of a proper finite difference approximation for image gradient computation, where the simple central finite difference approximation achieves superior performance compared to the Sobel filter and alternative gradient filters widely employed in image processing [12].

In this paper four more gradient filters are investigated in addition to the standard central difference scheme and Sobel kernel, where further information can be found in [3, 20].

$$\beta f'_{i-2} + \alpha f'_{i-1} + f'_i + \alpha f'_{i+1} + \beta f'_{i+2} = c\frac{f_{i+3} - f_{i-3}}{6} + b\frac{f_{i+2} - f_{i-2}}{4} + a\frac{f_{i+1} - f_{i-1}}{2}, \quad (2)$$

where the set of parameters $\{\alpha,\beta,a,b,c\}$ is defined either by

$$\left.\begin{aligned}\alpha = 0.5771439,\ \ \beta = 0.0896406\\ a = 1.302566,\ \ b = 0.99355,\ \ c = 0.03750245\end{aligned}\right\} \quad \text{Lele scheme} \tag{3}$$

as suggested by Lele [20], or by

$$\left.\alpha = \frac{3}{5},\ \ \beta = \frac{21}{200},\ \ a = \frac{63}{50},\ \ b = \frac{219}{200},\ \ c = \frac{7}{125}\right\} \quad \text{Fourier-Pade-Galerkin scheme} \tag{4}$$

as proposed in [8].

$$D_x = \frac{1}{2(w+2)}\begin{bmatrix}-1 & 0 & 1\\ w & 0 & w\\ -1 & 0 & 1\end{bmatrix} \tag{5}$$

where Bickley [4] and Scharr [17] set $w = 4$ and $w = 10/3$ respectively.

Both (2), (3) and (2), (4) deliver very accurate approximations of the 1st-order derivative not only for low frequencies but also for middle-range frequencies, while (5) produces superior orientations compared to magnitude values with the lowest computational expense. The schemes (2), (3) and (2), (4) are called implicit since they require solving a system of linear equations, so we call (5) explicit schemes.

4 Results

Figure 3: Weizmann Dataset showing frames from normal (left four) and deformation (right four) sequences.

Dataset

The Weizmann dataset contains three separate datasets (Figure 3). The first contains normal sequences and ten actions (run, walk, skip, jump, gallop sideways, one-hand wave, two-hand wave, bend, jump in place and jumping jack). The background is static permitting simple background subtraction. The remaining datasets are employed for robustness evaluation and contain varying viewpoints (0° to 81° in increments of 9°) and deformations (walk with a dog, swing a bag, wear a skirt, occluded feet, occluded by a pole, sleepwalk, walk with a limp, walk with knees up and walk with a briefcase). All datasets contain horizontally aligned silhouettes, therefore only height normalisation is required for GEI construction.

Action Class Decomposition

Considering GEI gradient distribution, strong gradients are restricted to dynamic areas, which for dynamic actions is widespread while limited to limb locations for static actions. Furthermore, dividing actions prior to classification promotes reduced processing time, and can be performed during pre-processing via silhouette analysis and a threshold based on

global translation. Performance of all actions versus static and dynamic actions is analysed. For the Weizmann dataset, static actions are defined as one-hand wave, two-hand wave, bend, jump in place and jumping jack, while dynamic actions are run, walk, skip, jump and gallop sideways.

SVM

SVMs [10] are a supervised learning technique widely employed for classification within computer vision and require a training and classification stage. Five kernels are employed to investigate performance: Linear, Quadratic, Polynomial, Gaussian Radial Basis Function and Multilayer Perceptron. Multi-class SVM is performed as one-versus-all (OVA) and one-versus-one (OVO) binary classification. OVA is a winner-takes-all approach which is heuristic to some degree [27], while OVO requires more classifiers than OVA leading to higher computational complexity. High dimensional feature vectors are not favourable for SVM with limited computational system [11]. Leave-one-out cross validation is applied for each sequence obtaining an average correct classification value enabling evaluation of HOG variables (gradient filter, cell and bin size) and performance against alternative techniques employing the Weizmann dataset.

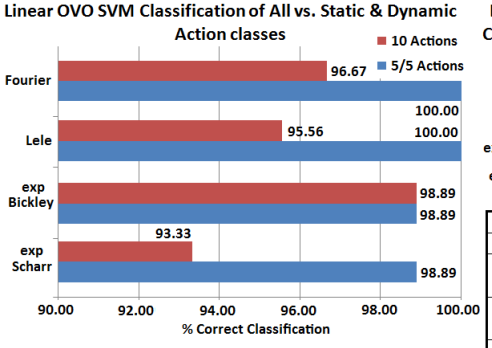
Weizmann Dataset Results

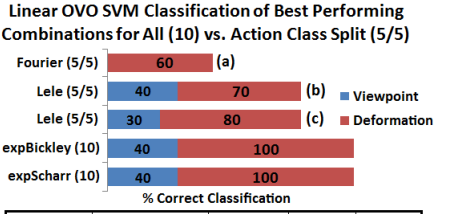
The HOG variables of interest are gradient filter and cell and bin sizes, where gradient filters vary in accuracy and computational complexity, while ten cell and bin sizes are investigated, where large values produce high dimensionality feature vectors. The first dataset, containing normal sequences, is employed to evaluate performance of all gradient filters and cell and bin sizes, where the best performing variables are evaluated on covariate factor sequences. While both OVA and OVO SVM results are investigated, the OVA results are poor in comparison to OVO, where similar results are observed by Hsu and Lin [16]. A linear kernel continually outperforms alternatives and indicates data is easily separable. The results for Linear OVO SVM can be seen in Figure 4, where normal and covariate factor sequences are shown on the left and right respectively for the top two performing gradient filters for all and static and dynamic actions, where associated cell and bin sizes are shown on the bottom right. While a number of gradient filters are employed for analysis, evaluation is performed on a single dataset and application to more datasets of varying complexity is required to avoid bias and achieve a more comprehensive evaluation of HOG parameters. While silhouette masks are normalised in direction of travel, analysis of GEIs with different directions [22] is required to determine robustness.

5 Discussion

Normal Action Recognition Results

The results for normal action recognition are shown on the left of Figure 4, where performance of static and dynamic actions is summed for comparison against that of all actions. The optimum parameters for all and static and dynamic actions are not shared, where static and dynamic actions outperforms all actions by 1.11%, which while a small margin, indicates the different parameter requirements likely due to gradient distribution in the GEI. Further decomposing the results, dynamic actions tend to outperform the static by 2.22%. The best performing gradient filters for all and static and dynamic actions are the explicit





Action	Gradient Filter	Bin Size	Cell Size	Total
5/5 Actions				**100%**
Static	Fourier-Pade-Galérikin	3	4,8,10	100
Dynamic		12	**(a)** 4	
Static	Lele	3	8,10	100
Dynamic		21	**(b)** 2,3 **(c)**	
10 Actions				**98.89%**
All Actions	exp Bickley	30	7	98.89
All Actions	exp Scharr	30	7	98.89

Figure 4: Results for action (left) and covariate factor recognition (right) with corresponding cell and bin sizes (bottom right)

Bickley and Scharr Schemes (100%) and Fourier-Padé-Galerikin and Lele (98.89%) respectively. Considering individual parameter requirements, all actions require a large bin and cell size, while the static and dynamic actions require small bin and large cell size and large bin and small cell size respectively. The difference in static and dynamic parameters further highlights the necessity for alternative parameters. As a general rule, small bin and cell sizes perform poorly, however the aforementioned parameters generally perform well over gradient filters and performance is generally high. The best combination of parameters is better chosen after evaluation under covariate factors as a trade off in decreased performance may be required for increased robustness, therefore these parameter combinations are evaluated under the viewpoint variation and deformations from the Weizmann dataset. While 100% performance is achieved when actions are split into static and dynamic classes, this performance is very unlikely in real world data, but promotes the ability of the technique of GEI, HOG and SVM for action recognition given good quality data.

Covariate Factor Recognition Results

The results for covariate factor action recognition are shown on the right of Figure 4. Given covariate factor sequences are based on walk, only dynamic actions are evaluated where the best performing HOG parameters are further evaluated. While it is desirable for high performance across viewpoint variance and deformation, the latter is of greater importance since viewpoint analysis will be performed during future work.

Superior robustness is seen for all compared to static and dynamic actions particularly during deformation sequences. For the dynamic actions, the Fourier-Padé-Galerikin filter achieves poorest performance in robustness to viewpoint variation and deformation, while Lele achieves superior robustness with 27° and 18° tolerance and 70% and 80% to viewpoint variance and deformation respectively cell size dependent. All actions shows 100% classification during deformation, and viewpoint variance matches Lele with 27° tolerance. Overall, dynamic action parameters appear more sensitive, where misclassification is caused by occlusion of the discriminative GEI areas (walk with dog) or normal pattern of walk not being performed (walk with knees up, limp and sleepwalk). The lack of robustness may be

due to the larger parameter sizes given smaller values promote a degree of invariance to local geometric and photometric transformation if smaller than the bin size [12]

Given similar results during normal action recognition, evaluation of all versus static and dynamic actions is therefore based on robustness. Static and dynamic actions achieve superior performance during normal sequences and faster classification at the cost of reduced robustness. A trade-off between performance, robustness and classification speed occurs. Overall, static and dynamic actions offer more in comparison to all actions through application of a Lele filter and a bin and cell size of 21 and 2 respectively. This conclusion requires further evaluation on alternative and more complex datasets to avoid bias to the Weizmann dataset.

6 Conclusion

This paper demonstrates the benefit of combining GEI, HOG and SVM for action recognition purposes where actions split into static and dynamic action classes are superior, and performance at 100% is encouraging, with a 27° viewpoint tolerance and robustness to covariate factors. Implementation of a global grid-based approach has shown a degree of invariance to occlusion, viewpoint and noise, however performance relies on the visibility of discriminative dynamic GEI areas and silhouette sequences portraying the basic shape of human and action. Despite a single feature approach, results are strong, and subsequent fusion [13, 29] with alternative features may permit increased robustness. In comparison to silhouette sequences, the GEI represents action performance as a single compact 2D image and noise within frames is acceptable given normalisation mitigates the effects, however the discriminative ability of the GEI is affected most by covariates targeting the shape.

Given 100% performance using static and dynamic actions, comparison is complex as little work performs the same action split. Despite this, performance of static and dynamic actions is equal to Gorelick et al. [8], while both all and static and dynamic actions outperform approaches such as [9, 22] where the former does not require any pre-processing which is advantageous. Furthermore, both sets of results rank highly compared to a recent collection of techniques [23] achieved with the Weizmann dataset. Results and trade-off indicate actions benefit from decomposition into static and dynamic classes due to GEI gradient distribution, where a Lele filter and bin and cell size of 21 and 2 respectively are best suited. Further analysis on datasets with varying complexity is required to avoid bias and achieve a more comprehensive evaluation of parameter behaviour. However the action class split demonstrates the requirement of different cell and bin sizes. The action is split is particularly beneficial for reducing computational complexity during classification.

Results promote the benefit of higher dimensionality feature vectors and this approach may benefit from dimensionality reduction. More complex datasets, such as the KTH actions dataset [28], are required to avoid bias and confirm parameter values. Action class division requires implementation. Multi-camera datasets will permit viewpoint analysis and an approach similar to Rudoy and Zelnik-Manor [26] may be performed. Experimentation with HOF [8] and fusion of HOG and HOF [30] may achieve increased performance and robustness against covariates affecting shape and performance of action, however this may be dataset dependent. An interesting alternative to combat the affects of covariate factors,

with the additional bonus of application in low light and night scenes, is employment of infrared imagery, where successful results have been achieved for both gait [31] and action [14] recognition.

Acknowledgements

We would like to thank the anonymous reviewers of this paper for their valuable and constructive comments.

References

[1] J.K. Aggarwal and M.S. Ryoo. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):1 – 43, April 2011.

[2] Khalid Bashir, Tao Xiang, and Shaogang Gong. Gait representation using flow fields. In *Proceedings of the British Machine Vision Conference*, pages 113.1–113.11, 2009.

[3] A. Belyaev. On implicit image derivatives and their applications. In *Proceedings of the British Machine Vision Conference*, pages 1 – 12, 2011.

[4] W.G. Bickley. Finite difference formulae for the square lattice. *The Quarterly Journal of Mechanics and Applied Mathematics*, 1(1):35 – 42, 1948.

[5] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri. Actions as space-time shapes. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1395 – 1402, October 2005.

[6] A. F. Bobick and J. W. Davis. The recognition of human movement using temporal templates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(3): 257 – 267, March 2001.

[7] L. Cao, M. Dikmen, Y. Fu, and T.S. Huang. Gender recognition from body. In *Proceedings of the 16th ACM International Conference on Multimedia*, MM '08, pages 725 – 728, New York, NY, USA, 2008. ACM.

[8] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1932 – 1939, June 2009.

[9] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1932 – 1939, June 2009.

[10] C. Cortes and V. Vapnik. Support-vector networks. In *Machine Learning*, pages 273 – 297, 1995.

[11] N. Dalal. *Finding people in images and videos*. PhD thesis, Institut National Polytechnique de Grenoble, 2006.

[12] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886 – 893, June 2005.

[13] N. Dalal, B. Triggs, and C. Schmid. Human detection using oriented histograms of flow and appearance. In *Proceedings of the 9th European conference on Computer Vision - Volume Part II*, ECCV'06, pages 428 – 441, 2006.

[14] J. Han and B. Bhanu. Human activity recognition in thermal infrared imagery. In *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, page 17, June 2005.

[15] J. Han and B. Bhanu. Individual recognition using gait energy image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(2):316– 322, February 2006.

[16] C-H. Hsu and C-J. Lin. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, 13(2):415 – 425, March 2002.

[17] B. Jähne, H. Scharr, and S. Körkel. Principles of filter design. In *Handbook of Computer Vision and Applications*, volume 2, Signal Processing and Applications, pages 125 – 151. Academic Press, 1999.

[18] I. Laptev and G. Mori. Statistical and structural recognition of human actions: European conference on computer vision tutorial, 2010.

[19] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 – 8, June 2008.

[20] S.K. Lele. Compact finite difference schemes with spectral-like resolution. *Journal of Computational Physics*, 103(1):16 – 42, November 1992.

[21] C. Lin and K. Wang. A behavior classification based on enhanced gait energy image. In *Networking and Digital Society (ICNDS), 2010 2nd International Conference on*, volume 2, pages 589 – 592, May 2010.

[22] H.-W. Lin, M.-C. Hu, and J.-L. Wu. Gait-based action recognition via accelerated minimum incremental coding length classifier. In *Advances in Multimedia Modeling*, volume 7131 of *Lecture Notes in Computer Science*, pages 266 – 276. Springer Berlin / Heidelberg, 2012.

[23] H. Liu, R. Feris, and M.-T. Sun. Benchmarking datasets for human activity recognition. In *Visual Analysis of Humans*, pages 411 – 427. 2011.

[24] O. Ludwig, D. Delgado, V. Goncalves, and U. Nunes. Trainable classifier-fusion schemes: An application to pedestrian detection. In *Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on*, pages 1 – 6, October 2009.

[25] R. Poppe. A survey on vision-based human action recognition. *Image Vision Comput.*, 28(6):976 – 990, June 2010.

[26] D. Rudoy and L. Zelnik-Manor. Viewpoint selection for human actions. *International Journal of Computer Vision*, pages 1 – 12, 2011.

[27] B. Scholkopf and A.J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2001.

[28] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: a local svm approach. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, pages 32 – 36, August 2004.

[29] B. Sun, J. Yan, and Y. Liu. Human gait recognition by integrating motion feature and shape feature. In *Multimedia Technology (ICMT), 2010 International Conference on*, pages 1 – 4, Octber 2010.

[30] H. Wang, M.M. Ullah, A. Klaser, I. Laptev, and C. Schmid. Evaluation of local spatio-temporal features for action recognition. In *Proceedings of the British Machine Vision Conference*, London, United Kingdom, September 2009.

[31] Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin. Infrared gait recognition based on wavelet transform and support vector machine. *Pattern Recognition*, 43(8):2904 – 2910, 2010.