International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 4 (2012) pp. 334–340

A Fast Statistical Approach for Human Activity Recognition

Samy Sadek [1], Ayoub Al-Hamadi [1], Bernd Michaelis [1] and Usama Sayed [2]

[1]Institute for Electronics, Signal Processing and Communications (IESK),
Otto-von-Guericke-University Magdeburg, Magdeburg, Germany
samy.bakheet@ovgu.de, samy.technik@yahoo.de

[2]Electrical Engineering Department,
Assiut University, Egypt

Abstract: **An essential part of any activity recognition system claiming be truly real-time is the ability to perform feature extraction in real-time. We present, in this paper, a quite simple and computationally tractable approach for real-time human activity recognition that is based on simple statistical features. These features are simple and relatively small, accordingly they are easy and fast to be calculated, and further form a relatively low-dimensional feature space in which classification can be carried out robustly. On the Weizmann publicly benchmark dataset, promising results (i.e. 97.8%) have been achieved, showing the effectiveness of the proposed approach compared to the-state-of-the-art. Furthermore, the approach is quite fast and thus can provide timing guarantees to real-time applications and embedded systems.**

Keywords: Activity recognition, motion analysis, statistical moments, video interpretation

I. Introduction

One of the most intriguing areas of research in the fields of pattern recognition and artificial intelligence is the automatic understanding of human activities in video sequences, which has been the center of interest of many researchers over the last two decades. In spite of the voluminous existing literature on the analysis and interpretation of human motion motivated by the rise of security concerns and increased ubiquity and affordability of digital media production equipment, research on human action and event recognition is still at the embryonic stage of development. Therefore much additional work remains to be done to address the ongoing challenges. It is clear that developing good algorithms for solving the problem of action recognition would yield huge potential for a large number of potential applications, e.g., human-computer interaction, video surveillance, gesture recognition, robot learning and control, etc. In fact, the non-rigid nature of human body and clothes in video sequences resulting from drastic illumination changes, changing in pose, and erratic motion patterns presents the grand challenge to human detection and action recognition [1]. In addition, while the real-time performance is a major concern in computer vision, especially for embedded computer vision systems, the majority of state-of-the-art action recognition systems often employ sophisticated feature extraction and/or learning techniques, creating a barrier to the real-time performance of these systems. Thus there is a possibility of a trade-off between accuracy/reliability and computational load.

In this paper, we propose a conceptually simple and computationally efficient framework to recognize human actions from video sequences. All the features extracted here are basically based on a set of difference images formed for example by successive subtraction of each preceding frame from each current one. The proposed method is evaluated using the popular Weizmann dataset. Experimental results show that our method not only effectively guarantees the real-time requirements required by real-time applications but also performs comparably to more computationally intensive and sophisticated methods in the literature.

The remainder of the paper proceeds as follows. Section II reviews the existing literature, while Section III describes the proposed action recognition method. Experimental results and a comparison with four widely quoted recent approaches are presented in Section IV. At last, in Section V, we conclude the paper and point out the future work.

II. Related Work

For the past two decades or so, a significant body of research literature has been contributed, proposing and/or investigating various methodologies for human activity recognition from video sequences. Human action can generally be recognized using various visual cues such as motion [2–4] and shape [5–8]. Scanning the literature, one notices that a large body of work in action recognition focuses on using interest points and local feature descriptors [1, 9–11]. The local features are extracted from the region around each keypoint. These features are then quantized to provide a discrete set of visual words before they are fed into the classification module. Another thread of research is concerned with analyzing patterns of motion to recognize human actions. For instance, in [3], periodic motions are detected and classified to recognize actions. Like us, some other researchers have opted to use both motion and shape cues. For exam-

ple, in [12], Bobick and Davis use temporal templates, including motion-energy images and motion-history images to recognize human movement. In [13] the authors detect the similarity between video segments using a space-time correlation model. While Rodriguez *et al.* in [14] present a template-based approach using a Maximum Average Correlation Height (MACH) filter to capture intra-class variabilities, Jhuang *et al.* in [15] perform actions recognition by building a neurobiological model using spatio-temporal gradient. Additionally in [16], actions are recognized by training different SVM classifiers on the local features of shape and optical flow. In parallel, a great deal of work focuses on modeling and understanding human motions by constructing elaborated temporal dynamic models [17, 18]. Finally, there is also a fertile and broadly influential area of research that uses generative topic models for modeling and recognizing action categories based on the so-called Bag-of-Words (BoW) model. The underlying concept of a BoW is that the video sequences are represented by counting the number of occurrences of descriptor prototypes, so-called visual words [19].

III. Proposed Methodology

In this section, we discuss our proposed methodology for real-time action recognition. Fig. 1 is a simplified block diagram illustrating the main components of the proposed recognition architecture, and how they interact with each other in order to achieve effective functionality of the whole system. As shown in the block diagram, a sequence of difference images is first constructed from successive frames of a video sequence by subtracting the current frame from the previous one. Then local features are extracted from the difference images based on a variety of shape moment descriptors. Since the global information of motion intuitively appears to be more relevant and appropriate to the current action recognition task, the final feature vectors fed into the SVM classifiers are constructed using both local and global features. In the following subsections, we discuss each module of the baseline architecture aforementioned in Fig. 1 in further detail, with a particular focus on the feature extraction module.

A. Preprocessing

For later successful feature extraction and classification, it is important to preprocess all video sequences to remove noisy, erroneous, and incomplete data, and to prepare the representative features that are suitable for knowledge generation. To wipe off noise and weaken image distortion, all frames of each action snippet are smoothed by Gaussian convolution with a kernel of size 3×3 and variance $\sigma = 0.5$. The following feature extraction module is basically based on the difference image of adjacent frames, which is a good cue for moving objects in the image (see Fig. 2). The difference image is first formed between successive frames of a given action snippet. This can be realized by simply subtracting the current frame from its immediate predecessor on a pixel-by-pixel basis. Then the absolute value of this difference is compared with a predetermined threshold. More formally

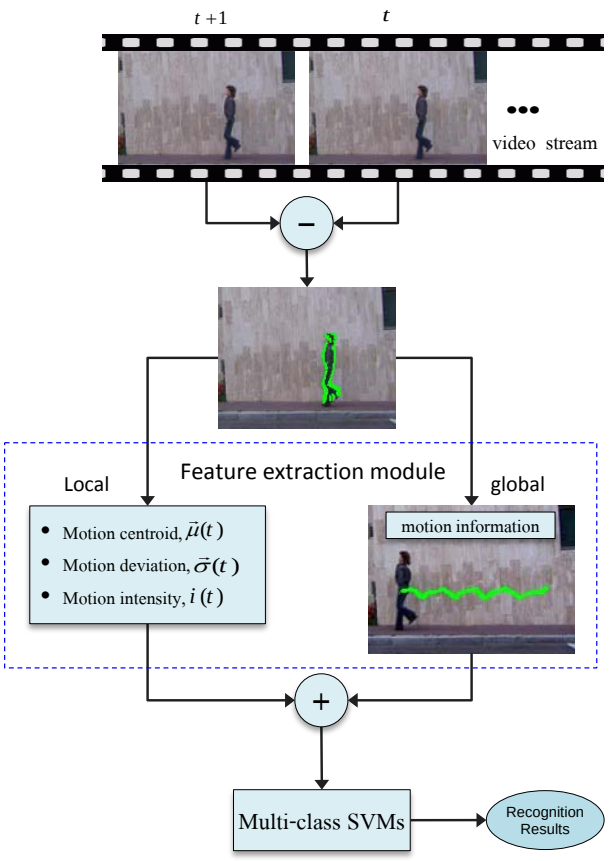

Figure. 1: Main structure of the proposed approach.

the difference image at a time t is given by:

$$\mathcal{D}_t(x, y) = |I_{t+1}(x, y) - I_t(x, y)| \tag{1}$$

where $I_t(x, y)$ and $I_{t+1}(x, y)$ are the frames at the time steps t and $t + 1$ respectively.

B. Feature extraction

Feature extraction is indeed the core of any recognition system, but is also the most challenging and time-consuming part. Further it was stated that the overall performance of the recognition system relies heavily on the feature extraction than the classification part. In particular, real-time feature extraction is a key component for any action recognition system that claims to be truly real-time. Many varieties of visual features can be used for human action recognition. In this work, the features that have been considered are derived from the difference images that primarily describe the shape of the moving human body parts. Such features represent a fundamental source of information regarding the interpretation of a specific human action. Furthermore the information of motion can be also extracted by following the trajectory of the motion centroid. The extracted features are primarily based on computing the moments of the difference images to specify the type of motion of a given action. Therefore the basic features are defined as:

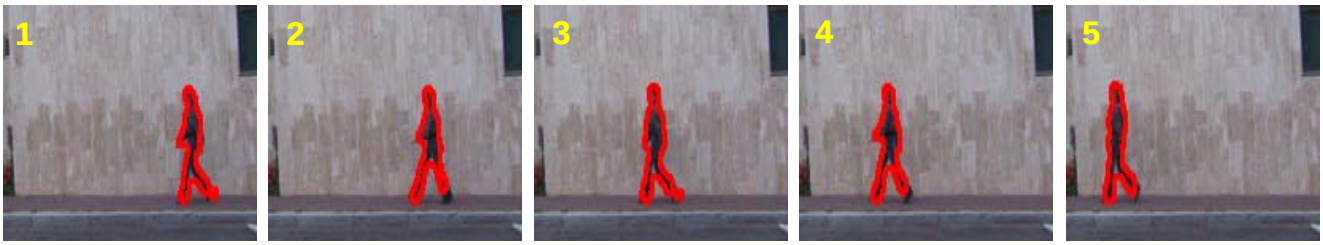

Figure. 2: Five shots from a "walking" action. The red-colored regions corresponding to the difference images, from where the features are extracted.

- center of motion $\vec{\mu}(t) = (\mu_x(t), \mu_y(t))$:

$$\mu_x(t) = \frac{\sum_{x,y} x\mathcal{D}_t(x,y)}{\sum_{x,y} \mathcal{D}_t(x,y)} \quad (2)$$

$$\mu_y(t) = \frac{\sum_{x,y} y\mathcal{D}_t(x,y)}{\sum_{x,y} \mathcal{D}_t(x,y)} \quad (3)$$

- mean absolute deviation from the center of motion $\vec{\sigma}(t) = (\sigma_x(t), \sigma_y(t))$:

$$\sigma_x(t) = \frac{\sum_{x,y}(x-\mu_x(t))\mathcal{D}_t(x,y)}{\sum_{x,y} \mathcal{D}_t(x,y)} \quad (4)$$

$$\sigma_y(t) = \frac{\sum_{x,y}(y-\mu_y(t))\mathcal{D}_t(x,y)}{\sum_{x,y} \mathcal{D}_t(x,y)} \quad (5)$$

- intensity of motion $i(t)$:

$$i(t) = \frac{\sum_{x,y} \mathcal{D}_t(x,y)}{\sum_{x,y} 1} \quad (6)$$

Thus the feature vector of a given action snippet at time t is given by

$$\vec{v}(t) = \big(\ \mu_x(t),\quad \mu_y(t),\quad \sigma_x(t),\quad \sigma_y(t),\quad i(t)\ \big)^T \quad (7)$$

Then all the feature vectors extracted from all frames within an action snippet are normalized to fit a zero-mean and a unit variance distribution. The normalized vectors obtained can now be used as shape contextual information for classification and matching. Many approaches in various object recognition applications directly combine these vectors to get one final vector per video and classify it using any classification algorithm. It would be worthwhile to note here that concatenating all the feature vectors extracted from all the frames of an action snippet will result in a large feature vector that might be less likely to be classified correctly, and not allow the system to run in real-time as intended. As a result, the effectiveness and efficiency of the the whole recognition system will be undermined or limited. To circumvent this problem and to reduce the dimensionality of the final feature vector of action snippet, first each action snippet is temporally divided into a number of overlapping time slices. Then all the feature vectors at a time-slice are weighted and averaged to obtain only one feature vector for each time-slice:

$$\vec{F_r} = \frac{1}{\tau}\sum_{t=0}^{\tau-1} \omega(t)\vec{v}(\frac{\tau}{2}r + t),\ r = 0, 1, 2, \ldots \quad (8)$$

where $\omega(t) = e^{-[\frac{(t-\frac{\tau}{2})}{2}]^2}$ is a weighting factor that defines a fuzzy membership for each feature vector within each time slice. Notably the closer the feature vector is to the center of the time-slice, the higher the weighting factor is. r is the time-slice index and τ is the length of each time-slice. All the feature vectors obtained from all the time-slices are eventually combined to generate one final vector that represents the feature descriptor for a specific action.

C. Global motion information

From the discussion in the previous section, it is seen that the local features obtained at each time-slice, are emphasized. Historically, global features have been successfully applied for automatic recognition in many applications of object recognition. This may permit and encourage us to combine the strengths of local and global features by fusing them to obtain robust and reliable recognition results. All the global information extracted here are based on calculating the motion centroid that delivers the center of motion. Therefore the temporal information that describe the distribution of motion are given by

$$\vec{\vartheta}_t = \frac{\vec{x_t} - \vec{x_0}}{t - t_0} \quad (9)$$

where $\vec{x_t}$ and $\vec{x_0}$ are the displacement vectors at times t and t_0 respectively. These features are proven to be most useful for our recognition task since they are very informative not only about the type of motion (e.g., translational or oscillatory), but also about the rate of motion (i.e., velocity). With these features, it would be able to distinguish, for example, between an action in which motion occurs over a relatively large area (e.g., running) and an action localized in a smaller region, where only small parts are in motion (e.g., one-hand waving). Hence, fusing local and global features provides a potential way by which notable improvements in the recognition performance can be achieved.

D. Action classification

In this section, we formulate the action recognition task as a multi-class learning problem, where there is one class for each action, and the goal is to assign an action to an individual in each video sequence. There are various supervised learning algorithms by which an action recognizer can be trained. Support Vector Machines (SVMs) are used in our framework due to their outstanding generalization capability and reputation of a highly accurate paradigm. SVMs [20] are based on the Structure Risk Minimization principle

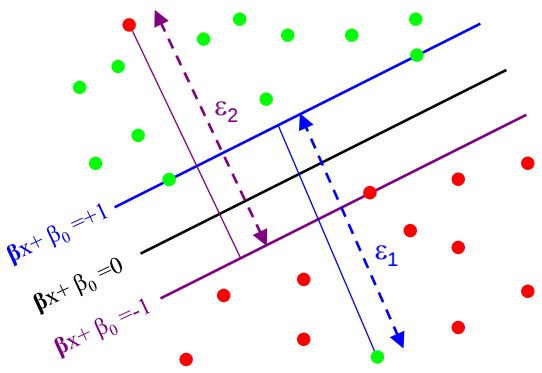

Figure. 3: Generalized optimal separating hyperplane.

from computational theory, and are a solution to data overfitting in neural networks. Originally, SVMs were designed to handle dichotomic classes in a higher dimensional space where a maximal separating hyperplane is created. On each side of this hyperplane, two parallel hyperplanes are conducted. Then SVM attempts to find the separating hyperplane that maximizes the distance between the two parallel hyperplanes. Intuitively, a good separation is achieved by the hyperplane having the largest distance (see Fig. 3). Hence, the larger the margin the lower the generalization error of the classifier. More formally, let $\mathcal{D} = \{(\mathbf{x}_i, y_i) \,|\, \mathbf{x}_i \in \mathbb{R}^d, y_i \in \{-1, +1\}\}$ be a training dataset, Coretes and Vapnik stated in their paper [20] that this problem are best addressed by allowing some examples to violate the margin constraints. These potential violations are formulated using some positive slack variables ξ_i and a penalty parameter $C \geq 0$ that penalize the margin violations. Thus the optimal separating hyperplane is determined by solving the following QP problem:

$$\min_{\boldsymbol{\beta}, \beta_0} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\sum_i \xi_i \tag{10}$$

subject to

$$(y_i(\langle \mathbf{x}_i, \boldsymbol{\beta}\rangle + \beta_0) \geq 1 - \xi_i \quad \forall i) \wedge (\xi_i \geq 0 \quad \forall i).$$

Geometrically, $\boldsymbol{\beta} \in \mathbb{R}^d$ is a vector going through the center and perpendicular to the separating hyperplane. The offset parameter β_0 is added to allow the margin to increase, and to not force the hyperplane to pass through the origin that restricts the solution. For computational purposes it is more convenient to solve SVM in its dual formulation. This can be accomplished by forming the Lagrangian and then optimizing over the Lagrange multiplier $\boldsymbol{\alpha}$. The resulting decision function has weight vector $\boldsymbol{\beta} = \sum_i \alpha_i \mathbf{x}_i y_i$, $0 \leq \alpha_i \leq C$. The instances $\mathbf{x}_i$ with $\alpha_i > 0$ are called *support vectors*, as they uniquely define the maximum margin hyperplane. In our approach, several classes of actions are created. Several one-vs-all SVM classifiers are trained using the features extracted from the action snippets in the training dataset. Both local features and global information of motion are combined to generate one feature vector per action snippet. All the feature vectors of trained sequences in the dataset are eventually fed into the SVM classifiers for the final decision.

IV. Experiments

In this section, the experiments conducted to show the performance of the proposed method are described. To assess the reliability of the method, the results obtained are compared with those reported in the literature for action recognition. All experiments were preformed on the popular Weizmann action dataset provided by Blank *et al.* [21] in 2005, which contains a total of 90 video clips (i.e., 5098 frames) performed by 9 individuals. Each video clip contains one person performing an action. There are 10 categories of action involved in the dataset, namely *walking*, *running*, *jumping*, *jumping-in-place*, *bending*, *jacking*, *skipping*, *galloping-sideways*, *one-hand-waving* and *two-hand-waving*. Typically, all the clips in the dataset are sampled at 25Hz and last about 2 seconds with image frame size of 180×144. A sample frame for each action in the Weizmann dataset is illustrated in Fig. 4. In order to provide an unbiased estimate of the generalization abilities of the proposed method, we used the leave-one-out cross-validation technique in the validation process. As the name suggests, this involves using a group of sequences from a single subject in the original dataset as the testing data, and the remaining sequences as the training data. This is repeated such that each group of sequences in the dataset is used once as the validation. More specifically, the sequences of 8 subjects were used for training and the sequences of the remaining subject were used for validation data. SVMs with Gaussian radial basis function kernel are trained on the training set, while the evaluation of the recognition performance is performed on the test set.
The recognition results obtained by the proposed method are summarized in a confusion matrix in Table 1, where correct responses define the main diagonal. From the figures in the matrix, a number of points can be drawn. The majority of actions are correctly classified. An average recognition rate of 97.8% is achieved with our proposed method. What is more, there is a clear distinction between arm actions and leg actions. The mistakes where confusions occur are only between *skip* and *jump* actions and between *jump* and *run* actions. This intuitively seems to be reasonable due to the fact of high closeness or similarity among the actions in each pair of these actions. In order to quantify the effectiveness of the proposed method, the results obtained are compared qualitatively with those obtained previously by other investigators. The outcome of this comparison is presented in Table 2. In light of this comparison, we can see that the proposed method is competitive with the state-of-the-art methods. It is worthwhile to mention that all the methods [10, 22–24] that we compared our method with, except the method proposed in [25], have used similar experimental setups, thus the comparison seems to be meaningful and fair. A final remark concerns the real-time performance of our approach. The proposed action recognizer runs at 18fps on average (using a 2.8 GHz Intel dual core machine with 4 GB of RAM, running 32-bit Windows 7 Professional). This clearly indicates that our recognition method is very amenable to working with real-time applications and embedded systems.

Figure. 4: A sample frame for each action in the Weizmann action dataset [21].

Table 1: Confusion matrix of the proposed method

ACTION	wave2	wave1	walk	skip	side	run	pjump	jump	jack	bend
wave2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
wave1	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
walk	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
skip	0.00	0.00	0.00	0.89	0.00	0.00	0.00	0.11	0.00	0.00
side	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
run	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
pjump	0.00	0.00	0.11	0.00	0.00	0.00	1.00	0.00	0.00	0.00
jump	0.00	0.00	0.03	0.00	0.00	0.11	0.00	0.89	0.00	0.00
jack	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
bend	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table 2: Comparison with four widely-quoted methods.

METHOD	ACCURACY
Proposed method	**97.8%**
Fathi *et al.* [25]	100%
Bregonzio *et al.* [22]	96.6%
Zhang *et al.* [23]	92.8%
Niebles *et al.* [24]	90.0%
Dollár *et al.* [10]	85.2%

V. Conclusion and Future Work

In this paper we have presented a computationally efficient method for real-time human action recognition using a finite set of features directly derived from difference frames of action snippet. Such features are very cheap to compute and form a relatively low dimensional feature space in which classification can be carried out robustly. Moreover partitioning action snippets into several time-slices in a fuzzy manner allows the model to be more robust to shape deformations and time wrapping effects. The results obtained are in a good agreement with those obtainable with much more sophisticated and computationally intensive methods in the literature. Furthermore the proposed method achieves real-time performance and thus can offer latency guarantees to real-time applications. However it would be advantageous to explore the empirical validation of the method on more complex realistic datasets presenting many technical challenges in data handling such as object articulation, occlusion, and significant background clutter. Such issues are of great interest and need to be tackled by our future work.

Acknowledgments

This work is supported by Forschungspraemie (BMBF-Förderung, FKZ: 03FPB00213) and Transregional Collaborative Research Centre SFB/TRR 62 Companion-Technology for Cognitive Technical Systems funded by the German Research Foundation (DFG).

References

[1] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, "Human action recognition: A novel scheme using fuzzy log-polar histogram and temporal self-similarity," *EURASIP Journal on Advances in Signal Processing*, January 2011.

[2] Y. G. Jiang, C. W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in *ACM Int. Conf. on Image and Video Retrieval*, vol. 8, pp. 494–501, 2007.

[3] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," *IEEE Transactions on PAMI*, vol. 22, pp. 781–796, August 2000.

[4] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, "An efficient method for real-time activity recognition," in *Proceedings of the International Conference on Soft Computing and Pattern Recognition (SoCPaR 2010)*, (Paris, France), pp. 7–10, December 2010.

[5] C. Thuran and V. Hlaváč, "Pose primitive based human action recognition in videos or still images," in *CVPR*, 2008.

[6] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, "Human activity recognition: A scheme using multiple cues," in *Proceedings of the International Symposium on Visual Computing (ISVC'10)*, vol. 1, (Las Vegas, Nevada, USA), pp. 574–583, November 2010.

[7] W.-L. Lu, K. Okuma, and J. J. Little, "Tracking and recognizing actions of multiple hockey players using the boosted particle filter," *Image and Vision Computing*, vol. 27, pp. 189–205, January 2009.

[8] S. Sadek, A. Al-Hamadi, M. Elmezain, B. Michaelis, and U. Sayed, "Human activity recognition using temporal shape moments," in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT2010)*, (Luxor, Egypt), pp. 79–84, December 2010.

[9] S. Sadek, A. Al-Hamadi, B. Michaelis, and U. Sayed, "Towards robust human action retrieval in video," in *Proceedings of the British Machine Vision Conference (BMVC'10)*, (Aberystwyth, UK), September 2010.

[10] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *IEEE International Workshop on VS-PETS*, 2005.

[11] J. Liu and M. Shah, "Learning human actions via information maximization," in *CVPR*, 2008.

[12] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on PAMI*, vol. 23, no. 3, pp. 257–267, 2001.

[13] E. Shechtman and M. Irani, "Space-time behavior based correlation," in *CVPR*, vol. 1, pp. 405–412, 2005.

[14] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH: A spatio-temporal maximum average correlation height filter for action recognition," in *CVPR*, 2008.

[15] H. Jhuang, T. Serre, L. Wolf, and T. Poggio, "A biologically inspired system for action recognition," in *ICCV*, pp. 257–267, 2007.

[16] K. Schindler and L. V. Gool, "Action snippets: How many frames does action recognition require?," in *CVPR*, 2008.

[17] B. Laxton, J. Lim, and D. Kriegman, "Leveraging temporal, contextual and ordering constraints for recognizing complex activities in video," in *CVPR*, pp. 1–8, 2007.

[18] N. Olivera, A. Garg, and E. Horvitz, "Layered representations for learning and inferring office activity from multiple sensory channels," *Computer Vision and Image Understanding*, vol. 96, no. 2, pp. 163–180, 2004.

[19] D. M. Blei and J. D. Lafferty, "Correlated topic models," *Advances in Neural Information Processing Systems (NIPS)*, vol. 18, pp. 147–154, 2006.

[20] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[21] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in *ICCV*, vol. 2, pp. 1395–1402, 2005.

[22] M. Bregonzio, S. Gong, and T. Xiang, "Recognising action as clouds of space-time interest points," in *CVPR*, 2009.

[23] Z. Zhang, Y. Hu, S. Chan, and L.-T. Chia, "Motion context: A new representation for human action recognition," in *ECCV*, vol. 4, pp. 817–829, 2008.

[24] J. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," *International Journal of Computer Vision*, vol. 79, no. 3, pp. 299–318, 2008.

[25] A. Fathi and G. Mori, "Action recognition by learning midlevel motion features," in *CVPR*, 2008.

Author Biographies

Samy Sadek received the B.Sc. and M.Sc. degrees in Computer Science from the University of Sohag, Sohag, Egypt in 2000, and 2005 respectively. He is currently working toward the Ph.D. degree at the Institute of Electronics, Signal Processing, and Communications (IESK), Otto-von-Guericke University Magdeburg, Magdeburg, Germany. His current research interests are geared towards high-level recognition problems in machine vision, such as human activity/event recognition, human pose estimation, object/scene recognition, etc. He also works on various multimedia related applications, such as visual surveillance, content-based image/video retrieval, video annotation.

Ayoub Al-Hamadi received his Masters Degree in Electrical Engineering and Information Technology in 1997 and his Ph.D. in Technical Computer Science at the Otto-von-Guericke-University of Magdeburg, Germany in 2001. Since 2003 he has been Junior-Research-Group-Leader at the Institute for Electronics, Signal Processing and Communications at the Otto-von-Guericke-University Magdeburg. In 2008 he became Professor of Neuro-Information Technology at the Otto-von-Guericke University Magdeburg. In May 2010, Prof. Al-Hamadi received the Habilitation in Artificial Intelligence and the Venia Legendi in the scientific field of "Pattern Recognition and Image Processing" from Otto-von-Guericke-University Magdeburg, Germany. He is the author of more than 150 articles in peer-reviewed international journals, conferences and books.

Bernd Michaelis received the degree of Diploma Engineer for Electronics from the Technische Hochschule Magdeburg in 1971. He received the first and second doctoral degree in 1974 and 1980 respectively. In 1980 he joined the Joint Institute for Nuclear Research Dubna. In 1984 he became Hochschuldozent at the Technische Hochschule Magdeburg. His research interests include Image processing, Artificial neural networks, Biological neural networks, Microcomputers and Processor architectures. Prof. Michaelis is the author

of more than 200 articles in peer-reviewed international journals and conferences.

Usama Sayed received his B.Sc. and M.Sc. degrees from Assiut University, in 1985 and 1993, respectively, and his Ph.D. degree from Czech Technical University in Prague, Czech Republic, in 2000. He authored and co-authored more than 85 scientific papers. Usama has been selected for the inclusion in 2010 Edition of the Marquis Who's Who in the World. His research interests include telecommunication technology, wireless technology, wireless Networks, RFID, image coding, speech coding, statistical signal processing, blind signal separation, and video coding.