International Journal of Computational Vision and Biomechanics
Vol. 3 No. 1 (January-June, 2017)

A Simple Approach for Low-Resolution Activity Recognition

Md. Atiqur Rahman Ahad, J. K. Tan, H. S. Kim and S. Ishikawa

Faculty of Engineering, Kyushu Institute of Technology, Japan

For video surveillance and other related purposes, we need to analyze and recognize actions from poor-quality and low-resolution video. However, human motion recognition from low-resolution video is very difficult task because due to low-resolution, much significant motion information and their correlations are missing from the image sequences. This paper exploited the appearance-based directional motion history image method to recognize various levels of video resolutions. This method can overcome the self-occlusion problem that arises from motion overwriting. Dataset having various levels of resolutions are employed to show the robustness of the method. A satisfactory result is achieved up to a limit after which the very low-resolution hinders the recognition process.

1. INTRODUCTION

Video surveillance, activity analysis, motion recognition, etc. are very crucial applications in computer vision. The demand and usage of CCTV is increasing for surveillance and hence low-resolution video analysis is crucial. However, low-resolution image simply consists of a small amount of pixels, causing the image to be jaggy. In various image and video processing, low-resolution processing is a demanding task in computer vision, and image processing arena (e.g., biometrics identification [1], range image sequence analysis in roads for safety [2], document analysis, face recognition [3], sports analysis [4], etc.). However, low-resolution video processing, motion analysis and recognition are very difficult due to the loss of significant motion information due to the presence of less number of pixels. For example, various techniques are developed for recognition of objects in photographs, but they often fail when applied to recognition of the same objects in video. A critical example of such a situation is seen in face recognition, where many technologies are already intensively used for passport verification and where there is no technology that can reliably identify a person from a surveillance video. The reason for this is that video provides images of much lower quality and resolution than that of photographs [5].

Directional motion history image (DMHI), a view-based template-matching method is presented to recognize various complex motions with satisfactory recognition result. The DMHI method is the extension to the basic motion history image (MHI) template method, developed by Bobick and Davis [6]. To recap what is the MHI method, we can state that the motion *history* image (MHI) can describe *how* the motion is moving in the image sequence. Another representation called motion *energy* image (MEI) can demonstrate the presence of any motion or a spatial pattern in the image sequence. Both MHI and MEI templates comprise the motion history image (MHI) template-matching method. The MHI is a scalar-valued image where more recently moving pixels are brighter, and vice versa, whereas, MEI is a binary image. The MHI $H_\tau(x, y, t)$ can be computed from update function $\Psi(x, y, t)$:

$$H_\tau(x,y,t) = \begin{cases} \tau & \text{if } \Psi(x,y,t) = 1 \\ \max(0, H_\tau(x,y,t-1) - \delta) & \text{otherwise} \end{cases} \quad (1)$$

Here, x, y and t show the position and time, $\Psi(x, y, t)$ signals object presence (or motion) in the current video image, τ decides the temporal duration of MHI (in terms of frames), and δ is the decay parameter. This update function is called for every new video frame analyzed in the sequence. The MEI is cumulative binary motion image, from which we can easily visualize the video sequence where it sweeps out a particular region of the image. We can generate the MEI by thresholding the MHI above zero:

$$E_\tau(x,y,t) = \begin{cases} 1 & \text{if } H_\tau(x,y,t) \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

This paper is organized as follows: Section 2 covers related works. Our DMHI method is discussed in Section 3. Section 4 covers experimental and classification procedures and feature vector calculation. Experimental results are analyzed in the next section. Finally, we concluded the paper in Section 6.

2. RELATED WORK ON MHI METHOD AND LOW-RESOLUTION PROCESSING

We covered some related works on motion history templates and low-resolution video processing approaches. This sub-section reviews advancements and applications based on the basic MHI approach. Various related surveys covered more details on action analysis and motion recognition [7-17]. In this paper, we used complex actions or activities. Using complex actions, we may face self-occlusion due to overwriting or overlapping from other motion parts. By self-occlusion, we mean that the recency of motion template in the MHI is not proper, because previous motion is deleted or overwritten by the later motion information, if the motion has overwriting in its original form [18-19]. For example, if a person sits down, and then the final MHI image should contain brighter pixels in the lower part of the image to represent the sitting down motion. And if a person stands up, then the upper part will be brighter due to the up motion in it. However, a 'sit-down' and then 'stand-up' motion will wipe up the 'sit-down' information, and only provide the later part of the motion (here, the 'stand-up') and thereby, this *self*-occlusion of the moving object or person overwrites the prior information. This is overwriting problem in the MHI approach and the basic MHI can not solve this problem [19]. In an update, the MHI is generalized by directly encoding the actual time in a floating point format which is called timed Motion History Image (tMHI) [20]. A gradient of the MHI is used to determine normal optical flow vector. Albu *et al.* [21] presents a new 3D motion representation, called the Volumetric Motion History Image, to be used for the analysis of irregularities in human actions. This work tried to overcome the limitation of the standard MHI related to motion self-occlusion and speed.

In MHI, at each pixel location, explicit information about its past is lost when current change are updated to the model with their corresponding MHI values 'jumping' to the maximal value. To overcome this problem, Pixel Signal Energy was introduced to measure the mean magnitude of pixel-level temporal energy over a period of time defined by a backward window [22]. This method suffers from its sensitivity to noise and is expensive to compute. Jin *et al.* [23] proposed an approach to generate motion-based pattern where moving objects are firstly segmented by employing the adaptive threshold-based change detection approach. Then they used scalar valued rear motion history image and front motion history image to represent how motion is evolved. This representation is then used to segment and measure the motion induced by the object. Later, the motion vectors with orientation and magnitude are generated from the chamfer distance. Finally, they derived one novel approach to generate intra motion history image for inside moving parts. As compared to MEI and MHI, Gait Energy Image [24] and Gait History Image [36] are developed for gait representation and recognition. GHI inherits the idea of MHI in the sense that the temporal and the spatial information can be recorded.

Pixel Change History (PCH) [25] measures the multi-scale temporal changes at each pixel. When no significant pixel-level visual change is detected at a particular location (x, y) in the current frame, pixel (x, y) will be treated as part of the background and the corresponding pixel change history starts to decay. Compressed domain human motion is recognized at the top of MHI approach by the introduction of the motion flow history (MFH) that quantifies the motion in compressed video domain [26]. Moreover, the MHI method is used for various interactive systems, such as interactive art demonstration, interactive play-space for kids [27], visual speech recognition [28], facial action unit (AU) analysis [19], visual surveillance and moving object tracking [29-31], threat assessment for automated visual surveillance [32], high-speed motion recognition [33-34], hand gesture recognition [35], gait analysis [24,36] unusual behavior detection and automatic event analysis [37], etc. Apart from above 2D approaches, few 3D extensions of the basic MHI method are available too [21,38-41]. These 3D approaches are view-invariant and are developed on the base of the MHI method [6].

Most of the works in human tracking and activity recognition are only appropriate for "near field" with higher resolution frames. Even the MHI [6] and the above variants require reasonably high-resolution video. Usually, detection of pedestrians is very difficult in surveillance applications, where the resolution of the images is very low (e.g., there may only be 100-200 pixels on the target) [43]. There are not much developments in low-resolution video analysis and action recognitions, though document analysis [44-45], license plate recognition [46], automatic detection and tracking of human hands [47] in low-resolution are available.

A neuro-associative approach to recognition which can both learn and identify an object from low-resolution low-quality video sequences is introduced by Gorodnichy [5]. This approach is derived from a mathematical model of biological visual memory, in which correlation-based projection learning is used to memorize a face from a video sequence and attractor-based association is performed to recognize a face over several video frames. In many sports, natural daily views or surveillance, the camera covers a large part of the sports arena or scene, so that the resolution of person's region is low. This makes the determination of the player's gestures and actions a challenging task, especially if there is large camera motion. To overcome these problems, Roh *et al.* [4,48] proposed a method based on curvature scale space (CSS) templates of the player's silhouette in low-resolution. The proposed spotting method provides probabilistic similarity and is robust to noisy sequences of data [48].

Jun *et al.* [49] proposed a new method of extracting four directional features for character recognition at low-resolution. The characteristic of their method is to use dynamic scenes. This is achieved by using images shifted by half a pixel in the low resolution. The recognition results showed effectiveness at low resolution. Another approach employed a new laser-based camera that produces reliable low-resolution depth images at video rates, to decompose and recognize hand poses. For low-resolution face recognition, Lee *et al.* [4] proposed a new method of extending the Support Vector Data Description learning methods for the one-class problem.

Efros *et al.* [50] recognized human actions at a distance, at resolutions where a whole person may be, say, 30 pixels tall. They introduced a motion descriptor based on optical flow measurements in a spatio-temporal volume for each stabilized human figure, and an associated similarity measure to be used in a nearest-neighbor framework. A pedestrian detection system that integrates image intensity information with motion information at very small scale is developed by [43]. They employed a detection style algorithm that scans a detector over two consecutive frames of a video sequence. The detector was trained using AdaBoost to take advantage of both motion and appearance information to detect a walking person. It detected pedestrians at very small scales (as small as 20x15 pixels). Cutler and Davis [53] developed a system that works directly on images which can be of low resolution and poor quality.

3. DMHI TEMPLATE REPRESENTATION

In this section, we present our directional motion history image (DMHI) template representation. Fig. 1 outlines the corresponding system-flow diagram for this method. In this case, consecutive frames of a video sequence are considered to recognize the activity of a person in the video sequence. Unlike the MHI method where background subtraction or frame differencing methods are employed to calculate the update function, this method computes gradient-based optical flow method between consecutive frames [54].

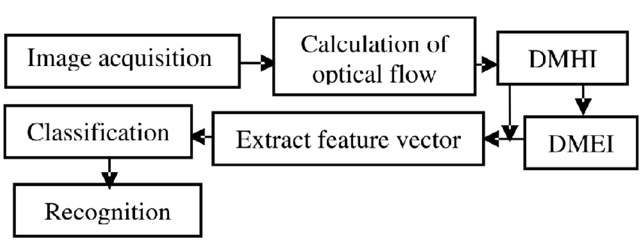

Figure 1: System Flow of this Recognition Approach

The optical flow vector is split into four different directions and then based on these four vectors, four different motion history images are computed (called directional motion history image, or DMHI) as shown in (3). In the similar fashion, directional motion energy images (DMEI) are computed after thresholding the motion energy templates above zero (as shown in Eq. 4).

Optical flow vector separates the upper, lower, left and right motions and thereby, the computed motion templates solely represent the directed components. Therefore, the overwriting problem can be solved. Based on a threshold th_p on pixel values, we compute four DMHIs.

$$H_\tau^{-x}(x,y,t)=\begin{cases}\tau & \text{if } \Psi_x^-(x,y,t)>th_p\\ \max(0,\ H_\tau^{-x}(x,y,t\text{-}1)\text{-}\delta) & \text{otherwise}\end{cases}$$

$$H_\tau^{+x}(x,y,t)=\begin{cases}\tau & \text{if } \Psi_x^+(x,y,t)>th_p\\ \max(0,\ H_\tau^{+x}(x,y,t\text{-}1)\text{-}\delta) & \text{otherwise}\end{cases}$$

$$H_\tau^{-y}(x,y,t)=\begin{cases}\tau & \text{if } \Psi_y^-(x,y,t)>th_p\\ \max(0,\ H_\tau^{-y}(x,y,t\text{-}1)\text{-}\delta) & \text{otherwise}\end{cases}$$

$$H_\tau^{+y}(x,y,t)=\begin{cases}\tau & \text{if } \Psi_y^+(x,y,t)>th_p\\ \max(0,\ H_\tau^{+y}(x,y,t\text{-}1)\text{-}\delta) & \text{otherwise}\end{cases}$$

$$E_\tau^{-x}(x,y,t)=\begin{cases}1 & \text{if } H_\tau^{-x}(x,y,t)>0\\ 0 & \text{otherwise}\end{cases}$$

$$E_\tau^{+x}(x,y,t)=\begin{cases}1 & \text{if } H_\tau^{+x}(x,y,t)>0\\ 0 & \text{otherwise}\end{cases}$$

$$E_\tau^{-y}(x,y,t)=\begin{cases}1 & \text{if } H_\tau^{-y}(x,y,t)>0\\ 0 & \text{otherwise}\end{cases} \quad (4)$$

$$E_\tau^{+y}(x,y,t)=\begin{cases}1 & \text{if } H_\tau^{+y}(x,y,t)>0\\ 0 & \text{otherwise}\end{cases}$$

Filtering is performed to smoothen these motion history images and to remove the salt-pepper noise. These directional components are normalized before calculating feature vectors.

4. EXPERIMENTAL PROCEDURES FOR CLASSIFICATION AND RECOGNITION

We compute moments of the final DMHIs and DMEIs to extract feature vectors for each motion, (as pointed in Fig .1). Geometrical moment invariants are widely used as global-based method. Hu [55] moments are widely used for various 2D pattern recognition applications [56]. Shape representation by statistical moments is a classical technique in the literature [57-58]. These moments are invariant to translation, scaling and rotation, though not invariant to changes in illumination [59]. 3D extension to the 2D moment invariants are introduced by [51]. As we compute the Hu moments in this experiment, the following parts of this section briefly sketch the generation of these moments. The 2D $(p + q)^{th}$ order central moments of a density distribution function $\rho(x, y)$ is defined as,

$$\mu_{pq}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\bar{x})^p(y-\bar{y})^q\rho(x,y)d(x-\bar{x})\ \ d(y-\bar{y}). \quad (5)$$

It is easy to express the central moments in terms of the ordinary moments. The first three central moments are constant. Hence, from the second and third order moments, we compute the following six absolute orthogonal invariants,

$$
\begin{aligned}
I_1 &= \mu_{20} + \mu_{02}, \\
I_2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2, \\
I_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2, \\
I_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2, \\
I_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + \\
&\quad (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2], \\
I_6 &= (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}),
\end{aligned}
\tag{6}
$$

From these higher order invariants we can realize that higher order invariants are more difficult to derive and quite complex to compute. Hence, these invariants (I_1 to I_7) are considered and higher than these are ignored due to their increased complexities and more tendencies towards noise [60-61]. For the DMHI approach, we initially computed normalized 0th order moment for each template ($\overline{m_{00}^{\varpi}}$ where $\varpi \in$ {*up, down, left, right*}). We estimated eight normalized 0th order moments for the eight templates (DMHIs and DMEIs). Hence, we get a 64-dimensional feature vector (7 Hu moments for each of the eight templates for DMHIs and DMEIs, in addition to their normalized 0th order moments) for a motion. In our analysis, we find that the presence of 0th order moments increases the robustness and recognition result because 0th order moment provides the total mass of an image. These normalized 0th order invariants confer the relative mass of the motion area in the scene. We can define FV_{dmhi} as:

$$FV_{dmhi} = (\forall h)(I_{1\to 7}) \cup (\forall e)(I_{1\to 7}) \cup (\forall h)\left(\overline{m_{00}^{\varpi}}\right) \cup (\forall e)\left(\overline{m_{00}^{\varpi}}\right). \tag{8}$$

where, h denotes the motion *h*istory image components, and e stands for *e*nergy components of the DMHI method.

We used ten various complex aerobics for recognition in this experiment. These are done by eight subjects irrespective of size, height and dress. No special marker or arrangement was considered. These video were taken using a frontal-view digital video camera with almost same illumination in indoor. Few frames are shown in Fig. 2. For classification, k-nearest-neighbor classifier is employed. Leave-one-out cross-validation partitioning scheme is considered. It means out of $\boldsymbol{N}$ samples from each of the c classes per database, $\boldsymbol{N} - 1$ of them are used

Figure 2: Few Frames for the Ten Aerobics Employed in this Experiment

to train (design) the classifier and the remaining one to test it [60]. This process is repeated ***N*** times, each time leaving a different sample out. Therefore, all of the samples are ultimately used for testing. This process is repeated and the resultant recognition rate is averaged. Based on the above-mentioned feature vectors and classification methods, the recognition results are achieved.

5. RESULT AND ANALYSIS

This section presents the recognition result and analyses of our DMHI method for various activities with various resolution datasets. Our key concern here is to demonstrate the performance of this DMHI method using videos of various low-resolution levels. In this experiment, we used ten different complex activities. To show the performance of the DMHI method, we compared it with the basic MHI method. We achieved far better results with the DMHI method. Table I shows the comparative results with 10 actions. We used k=1 and k=3 in the k-nearest-neighbor framework. As evident, the dataset is very complex in nature, and each activity contains motions in almost all four directions – left, right, up and down. So the basic MHI showed poor recognition result due to the presence of self-occlusion in the actions. In this case, the resolution of each frame was 320 × 240 pixels. Later for low-resolution motion recognition, we reduced the image resolution: from half (i.e., small email size: 160 × 120 pixels) to one-tenth (32 × 24 pixels) of the original frame size (320 × 240 pixels), though Wu *et al.* [46] considered 320 × 240 pixels as very low resolution for license plat recognition. However, for human activity recognition and analysis from low resolution video is still very difficult task.

Fig. 3 shows the same frame with various resolutions. As obvious from Fig. 3, for 32x24 resolution, the information is very noisy and edges are less evident, and hence the computation of optical flow or other method is difficult for measuring proper motion information. Therefore, low-resolution image processing and motion recognition are difficult tasks, because due to low-resolution, we miss some important image information and their respective correlations.

Fig. 4 shows four energy image components for the first action, having five different resolution levels. It clearly depicts that with low-resolution, loss of motion information is evident. Note the last row (for 32x24 pixels) and others. If we compare with the DMEIs for 320x240, we can notice the differences. Hence, motion recognition with very low-resolution video is very difficult and sometimes deemed to be impossible. Table II shows the recognition rates of the first six actions, having different resolution rates for DMHI. The results demonstrate that this method can achieve good recognition results from low-resolution video sequences. In this table, we see that lower the resolution, poorer the recognition results except for 96×72 pixel, where we achieved better result. Based on our analysis, this result does not demonstrate that this low-resolution is the optimum for having good recognition results. For 80x60 or lower, we fail to get good and reasonable recognition results. This is due to the fact that we have less information in very low-resolution image frames. In this Table, we do not present the results for 32x24 pixel. This can be understood from the images of bottom-row of Fig. 4. We do not get enough information from those too low-resolution images and hence can not get reasonable results for 32x24 pixel. Therefore, we consider that our method can perform reasonably up to a limit of resolution and after that we can not achieve results.

We employed MHI too with low-resolution datasets; but it was found that the recognition rate for MHI was very poor (though we can predict that for simple actions, MHI should perform well in lower resolution too. Nevertheless, we did not perform any experiment with

Table I
Comparative Recognition Results between MHI and DMHI Methods

No.	*MHI (k=1)*	*DMHI (k=1)*	*MHI (k=3)*	*DMHI (k=3)*
1	50	100	50	100
2	00	100	00	100
3	87.5	100	87.5	100
4	50	100	50	100
5	00	87.5	12.5	87.5
6	25	87.5	25	87.5
7	50	87.5	50	87.5
8	50	87.5	50	75
9	100	87.5	100	87.5
10	50	100	37.5	100
Avg.	46.3	93.8	46.3	92.5

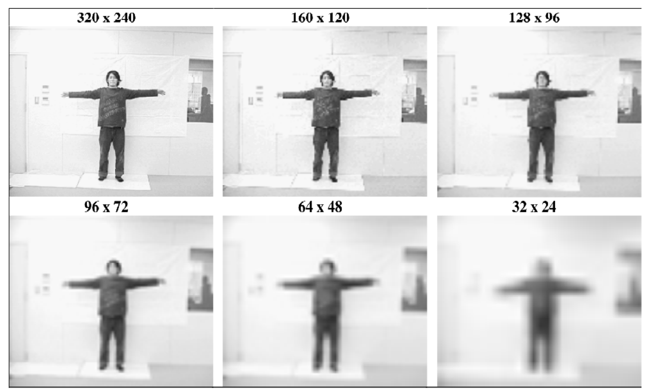

Figure 3: One Frame is Shown Here with Six Different Resolutions. Last Frame is Difficult to Process

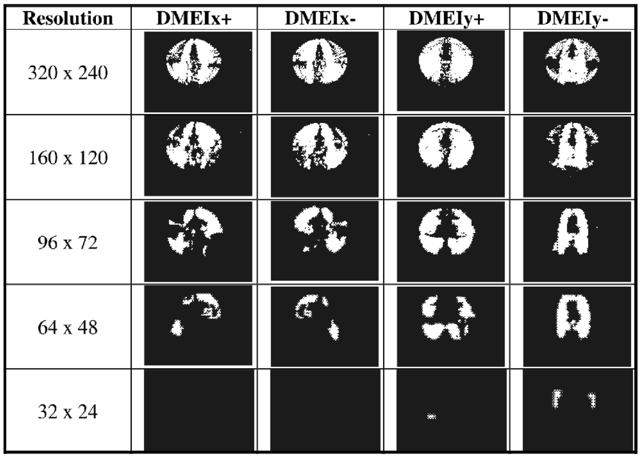

Figure 4: Four DMEIs for the First Action with Different Resolutions

Table II
Recognition With Various Resolutions

Frame Resolution	*Reduced size from original*	*Rec. Rate (%)* *K1*	*K3*
320 × 240	100%	95.9	95.9
160 × 120	50%	94.4	94.4
128 × 96	40%	89	83.4
96 × 72	30%	94.4	89
80 × 60	25%	75	78
64 × 48	20%	73	75

simple actions having low-resolution frames for the MHI method). Using the DMHI representation, we can easily achieve good recognition rate even though the resolution is low. As we have eight templates for history and energy images, with low-resolution video, we still can achieve overall good information about the action.

Table III
Recognition with Various Resolutions and Actions

No. of Actions	*320 × 240*		*160 × 120*		*64 × 48*	
	K1	*K3*	*K1*	*K3*	*K1*	*K3*
7	94.7	94.7	89.3	91.1	71.5	73.3
8	93.8	92.2	89.1	90.1	64.1	67.2
9	93.1	91.7	90.1	91.7	58.4	62.5
10	93.8	92.5	91.3	92.5	61.3	63.8

In Table III, results with different datasets are presented. We can notice from this table that average recognition rate is satisfactory and about 90% or more with various datasets. However, for video of very poor resolution (here, 64 × 48), we noticed deterioration in recognition rate for more variations in actions. Based on our analysis, we can admit that due to low-resolution, we missed important information in motion. Also, in this experiment, the background was a bit cluttered, behind the subject, sometimes passers-by passed through the corridor and these unwanted motions caused some added outliers. This experiment was taken inside a room with almost constant light illumination, and the dress, size, height, etc. of the subjects were diverse and these were not problem in this analysis.

6. CONCLUSIONS

As pointed above, motion recognition from low-resolution video sequences is very difficult, mainly due to the fact that low-resolution images are noisy and the image edges are less prevalent. In this paper, we demonstrated recognition of low-resolution motion datasets with a robust view-based template-matching motion recognition method called DMHI. This method is robust in low-resolution video data though very poor-quality video analysis and recognition seems to be difficult at this stage. For low resolution datasets, we achieved satisfactory recognition rate. This is very encouraging result with our DMHI method, because low resolution means lesser pixels to calculate and evaluate, and hence faster recognition. Low-resolution processing means less number of pixels to calculate and process; and in that way, it becomes faster to process. However, as evident from Fig. 4, we failed to get any significant information for very low-resolution, e.g., 32×24 resolution. Therefore, we can not achieve recognition results for every activity in very low-resolution video sequences. Therefore, we need to work to solve this problem in future. Moreover, to select the threshold automatically on-the-fly is very important but challenging task for this purpose. These issues can be addressed in future. Finally, we feel that this present concept can be useful for motion analysis and motion understanding in video, rather full recognition.

REFERENCES

[1] W. K. Kong and D. Zhang (2002), "Palmprint Texture Analysis Based on Low-resolution Images for Personal Authentication", *Int. Conf. on Pattern Recognition*, 3: 807-810.

[2] K. Sobottka (2000), "Analysis of Low-resolution Range Image Sequences", Ph.D. Thesis, 250 pages.

[3] M. C. Roh, W. J. Christmas, J. Kittler, and S. W. Lee (2006), "Robust Player Gesture Spotting and Recognition in Low-resolution Sports Video", Computer Vision – ECCV 2006, LNCS, 3954: 347-358.

[4] S. W. Lee, J. Park, and S. W. Lee (2006), "Low Resolution Face Recognition Based on Support Vector Data Description", *Pattern Recognition*, 39: 1809-1812.

[5] D. O. Gorodnichy (2005), "Associative Neural Networks as Means for Low-resolution Video-based Recognition", Int. Joint Conf. on Neural Networks (IJCNN'05), Canada, 6 pages.

[6] A. Bobick and J. Davis (2001), "The Recognition of Human Movement using Temporal Templates", *IEEE Trans. on PAMI*, 23(3): 257-267.

[7] Md. Atiqur Rahman Ahad, J. K. Tan, H. S. Kim, S. Ishikawa (2008), "Human Activity Recognition: Various Paradigms", Int. Conf. on Control, Automation and Systems (ICCAS), pp. 1896-1901.

[8] D. Gavrilla (1999), "The Visual Analysis of Human Movement: A Survey", CVIU, 73: 82-98.

[9] R. Wren, B. P. Clarkson, and A. P. Pentland (1999), "Understanding Purposeful Human Motion", *IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pp. 19-25.

[10] J. K. Aggarwal and Q. Cai (1999), "Human Motion Analysis: A Review", CVIU, 73: 428-440.

[11] J. LaViola (1999), "A Survey of Hand Posture and Gesture Recognition Techniques and Technology", Tech. Report CS-99-11, Brown University, Department of Computer Science, Providence RI, 80 pages.

[12] T. B. Moeslund and E. Granum (2001), "A Survey of Computer Vision-based Human Motion Capture", CVIU, 81: 231-268.

[13] J. J. L. Wang and S. Singh, "Video Analysis of Human Dynamics – A Survey," *Real-Time Imaging*, 9(5): 321-346.

[14] M. Pantic, A. Pentland, A. Nijholt, and T. S. Hunag (2006), "Human Computing and Machine Understanding of Human Behavior: A Survey", 8th Int. Conf. on Multimodal Interfaces, pp. 239-248.

[15] T. B. Moeslund, A. Hilton, and V. Kruger (2006), "A Survey of Advances in Vision-based Human Motion Capture and Analysis", CVIU, 104: 90-126.

[16] R. Poppe (2007), "Vision-based Human Motion Analysis: An Overview", CVIU, 108(1-2): 4-18.

[17] A. Jaimes and N. Sebe (2007), "Multimodal Human-computer Interaction: A Survey", CVIU, 108(1-2): 116-134.

[18] Md. Atiqur Rahman Ahad, T. Ogata, J. K. Tan, H. S. Kim, S. Ishikawa (2008), "Motion Recognition Approach to Solve Overwriting in Complex Actions", IEEE Int. Conf. Automatic Face and Gesture Recognition (FG'08), Amsterdam, The Netherland.

[19] M. Valstar, M. Pantic, and I. Patras (2004), "Motion History for Facial Action Detection in Video", *IEEE Int'l Conf. on Systems, Man and Cybernetics*, 1: 635-640.

[20] G. Bradski and J. Davis (2002), "Motion Segmentation and Pose Recognition with Motion History Gradients", *Mach. Vision Appl.*, 13(3): 174-184.

[21] A. Albu, B. Trevor, V. Naznin, and C. Beach (2007), "Analysis of Irregularities in Human Actions with Volumetric Motion History Images", *IEEE Workshop on Motion and Video Computing*, p. 16-16.

[22] J. Ng and S. Gong (2001), "Learning Pixel-wise Signal Energy for Understanding Semantics", *British Machine Vision Conference*, pp. 695-704

[23] T. Jin, M. K. H. Leung, and L. Li (2004), "Temporal Human Body Segmentation", Visualization, Imaging, and Image Processing, ed. Juan J. Villanieva, Acta Press, Marbella, Spain, ISSN: 1482-7921.

[24] J. Han, B. Bhanu (2006), "Individual Recognition using Gait Energy Image", *IEEE Trans. on PAMI*, 28(2): 316-322.

[25] T. Xiang and S. Gong (2006), "Beyond Tracking: Modelling Activity and Understanding Behaviour", *Int. J. Computer Vision*, 67(1): 21-51.

[26] R. Babu and K. Ramakrishnan (2003), "Compressed Domain Human Motion Recognition using Motion History Information", ICIP, 2: 321-324.

[27] A. Bobick, S. Intille, J. Davis, F. Baird, C. Pinhanez, L. Campbell, Y. Ivanov, A. Schutte, and A. Wilson (1999), "The Kidsroom: A Perceptually-based Interactive and Immersive Story Environment", *Presence: Teleoperators and Virtual Environmen.*, 8(4): 367-391.

[28] W. C. Yau, D. K. Kumar, S. P. Arjunan, and S. Kumar (2006), "Visual Speech Recognition using Image Moments and Multiresolution Wavelet", *Int. Conf. on Comp. Graphics, Imaging and Visualization*, pp. 194-199.

[29] Z. Yin and R. Collins (2006), "Moving Object Localization in Thermal Imagery by Forward-backward MHI", CVPR Workshop, pp. 133-140.

[30] L. Li, Q. Zeng, Y. Jiang, and H. Xia (2006), "Spatio-temporal Motion Segmentation and Tracking under Realistic Condition", Int. Symposium on Systems and Control in Aerospace and Astronautics, pp. 229-232.

[31] J. Piater and J. Crowley (2001), "Multi-modal Tracking of Interacting Targets using Gaussian Approximation", *IEEE Workshop on Performance Evaluation of Tracking and Surveillance*, pp. 141-147.

[32] T. Jan (2004), "Neural Network based Threat Assessment for Automated Visual Surveillance", *IEEE Int. Conf. on Neural Networks*, 2: 1309–1312.

[33] T. Ogata, J. K. Tan, and S. Ishikawa (2006), "High-speed Human Motion Recognition Based on a Motion History Image and an Eigenspace", *IEICE Trans. Information & Systems*, E89-D(1): 281-289.

[34] R. Rosales and S. Sclaroff (2003), "A Framework for Heading-Guided Recognition of Human Activity", CVIU, 91(3): 335-367.

[35] S. Kumar, D. Kumar, A. Sharma, and N. McLachlan (2003), "Classification of Hand Movements using Motion Templates and Geometrical Based Moments", *Int'l Conf. on Intelligent Sensing and Information Processing*, pp. 299-304.

[36] J. Liu, and N. Zhang (2007), "Gait History Image: A Novel Temporal Template for Gait Recognition", *IEEE ICME*, pp. 663-666.

[37] I. Petrás, C. Beleznai, Y. Dedeoğlu, M. Pardàs, *et al.* (2007), "Flexible Test-bed for Unusual Behavior Detection", *ACM Int. Conf. on Image and Video Retrieval*, pp. 105-108.

[38] H. K. Shin, S. W. Lee, S. W. Lee (2005), "Real-time Gesture Recognition using 3D Motion History Model", ICIC 2005, Part I, pp. 888–898.

[39] D. Weinland, R. Ronfard, and E. Boyer (2006), "Free Viewpoint Action Recognition using Motion History Volumes", *CVIU*, 104(2): 249-257.

[40] C. Canton-Ferrer, J. R. Casas, and M. Pardas (2006), "Human Model and Motion based 3D Action Recognition in Multiple View Scenarios", *Conf. on 14th European Signal Processing*, Italy, pp. 1-5.

[41] M. C. Roh, H. K. Shin, S. W. Lee, and S. W. Lee (2006), "Volume Motion Template for View-invariant Gesture Recognition", *ICPR*, 2: 1229-1232.

[42] Md. Atiqur Rahman Ahad, T. Ogata, J. K. Tan, H. S. Kim, S. Ishikawa (2008), "Directional Motion History Templates for Low Resolution Motion Recognition", 34th Annual Conf. IEEE Industrial Electronics Society (IECON), pp. 1875-1880.

[43] P. Viola, M. J. Jones, and D. Snow (2003), "Detecting Pedestrians using Patterns of Motion and Appearance", *Int. Conf. on Computer Vision*, 2: 734-742.

[44] S. Yanadume, Y. Mekada, I. Ide, and H. Murase (2004), "Recognition of Very low-resolution Characters from Motion

Images Captured by a Portable Digital Camera", Advances in Multimedia Information Processing - PCM 2004, pp. 247-254.

[45] M. Nomura, K. Yamamoto, H. Ohta, and K. Kato (2005), "A Proposal of the Effective Recognition Method for Low-resolution Characters from Motion Images", 8^{th} *Int. Conf. on Document Analysis and Recognition*, 2: 720-724.

[46] H.-H.P. Wu, H. H. Chen, R. J. Wu, and D. F. Shen (2006), "License Plate Extraction in Low Resolution Video", ICPR, 1: 824-827.

[47] N. B. Bo, M. N. Dailey, and B. Uyyanonvara (2007), "Robust Hand Tracking in Low-resolution Video Sequences", 3^{rd} IASTED Int. Conf. on Advances in Computer Science and Technology, pp. 228-233.

[48] M. C. Roh, W. J. Christmas, J. Kittler, and S. W. Lee (2008), "Gesture Spotting for Low-resolution Sports Video Annotation", *Pattern Recognition*, 41(3): 1124-1137.

[49] K. Jun, K. Kunihito, and Y. Kazuhiko (1999), "Character Recognition at Low Resolution with Video Camera", *Journal of the Institute of Image Information and Television Engineers*, 53(6): 867-872.

[50] A. A. Efros, A. C. Berg, G. Mori, and J. Malik (2003), "Recognizing Action at a Distance", *Int. Conf. on Computer Vision*, 2: 726-733.

[51] C. Lo and H. Don (1990), "Pattern Recognition using 3-D Moments", *Int. Conf. on Pattern Rec.*, 1: 540-544.

[52] Z. Mo and U. Neumann (2006), "Real-time Hand Pose Recognition using Low-resolution Depth Images", *IEEE Conf. on CVPR*, 2: 1499-1505.

[53] R. Cutler and L. Davis (2000), "Robust Real-time Periodic Motion Detection: *Analysis and applications", IEEE Trans. on PAMI*, 22: 781-796.

[54] W. J. Christmas (1998), "Spatial Filtering Requirements for Gradient-based Optical Flow Measurement", *British Machine Vision Conference*, pp. 185-194.

[55] M. K. Hu (1962), "Visual Pattern Recognition by Moment Invariants", *IRE Trans. on Info. Theory*, pp. 179-187.

[56] S. Zhang and G. Lu (2004), "Review of Shape Representation and Description Techniques", *Pattern Recognition*, 37: 1-19.

[57] B. Boulay, F. Bre´mond, and M. Thonnat (2006), "Applying 3D Human Model in a Posture Recognition System" *Pattern Recognition Letters*, 27: 1788-1796.

[58] D. Son, T. Dinh, V. Nam, T. Hanh, and H. Lam (2005), "Detection and Localization of Road Area in Traffic Video Sequences using Motion Information and Fuzzy-shadowed Sets", *Int. Symp. on Multimedia*, pp. 725-732.

[59] T. H. Reiss (1991), "The Revised Fundamental Theorem of Moment Invariants", *IEEE Trans. on PAMI*, 13(8): 830-834.

[60] A. Khotanzad and Y. H. Hong (1990), "Invariant Image Recognition by Zernike Moments", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(5): 489-497.

[61] Q. Chen, E. Petriu, and X. Yang (2004), "A Comparative Study of Fourier Descriptors and Hu's Seven Moment Invariants for Image Recognition", Canadian Conf. on Electrical and Computer Engineering, pp. 0103-0106.