ICROS-SICE International Joint Conference 2009
August 18-21, 2009, Fukuoka International Congress Center, Japan

Human Activity Analysis: Concentrating on Motion History Image and Its Variants

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

Department of Control Engineering, Kyushu Institute of Technology, Fukuoka, Japan
E-mail: atiqahad@ss10.cntl.kyutech.ac.jp

Abstract: The Motion History Image (MHI) method is a view-based temporal template approach, which is simple and robust in representing movements, and hence MHI is widely employed by various research groups for action recognition, motion analysis, etc. This paper overviews motion history image (MHI)-based human motion recognition techniques. Since the inception of the motion history image template for motion recognition, various progresses have been adopted to improve this basic MHI technique. We have presented almost all the important variants of the MHI method in this paper. This paper has pointed some areas for further research at the top of this method.

Keywords: Motion analysis, MHI, MEI, moment, template-based approach

1. INTRODUCTION

We present the works on human motion and behavior analysis based on the MHI and its variants for various applications. Various surveys [1-2] cited MHI method [3] as one of the important methods. MHI is a view-based or appearance-based template-matching approach. It is a scalar-valued image where more recently moving pixels are brighter, and vice versa. It expresses the motion flow or sequence by using the intensity of every pixel in temporal manner. We organize the paper as follows: Section 2 introduces basic MHI. Section 3 sums up the variants of MHI method. Section 4 presents applications based on these approaches. Discussions are presented in Section 5. We conclude in the paper Section 6.

2. MHI TEMPLATES

To describe the motion-shape and spatial distribution of motion, they proposed binary Motion Energy Image (MEI). And MHI image represents the recency of human action in a motion sequence. In fact, in order to describe *how* the motion is moving in the image sequence, one can form a motion-history image, and to represent *where* the motion or a spatial pattern is, one can demonstrate this by creating motion energy image. One of the advantages of MHI representations is that a range of times from frame to frame to several seconds may be encoded in a single frame, and in this way, MHIs span the time scale of human gestures. MHI $H_\tau(x,y,t)$ is computed from:

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

Here, x, y and t show the position and time, $\Psi(x,y,t)$ is update function that signals object's presence (or motion) in the current video image, the duration τ decides the temporal extent of the movement, and δ is the decay parameter. $\Psi(x,y,t)$ is called for every new video frame analyzed in the sequence. MEI is the cumulative binary motion image, which can describe where a motion is in the video sequence, computed from the start frame to the final frame. MEI is achieved by,

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

Both MHI and MEI images are important for representing motion information. Fig. 1 shows a flow diagram for recognizing actions using the MHI method. Fig. 2 shows some examples of MHI and MEI. First four columns are some sequential frames; images in 5th column are the corresponding MHIs; and images in the right-most column show the MEI images for both actions ('hand-waving and body-bending': upwards using left-hand (top row) and downwards using right-hand (bottom row)).

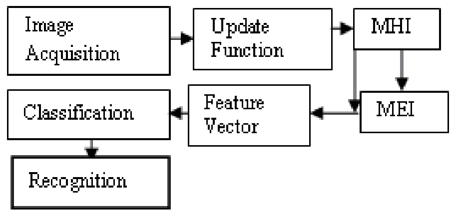

Figure 1. Basic flow diagram of the MHI method.

Figure 2. Example of MHI and MEI templates.

Fig. 3 shows the dependency of δ in calculating MHI templates. Changing in δ can give different information and these can be effective based on the dataset. The top-row shows final MHI images for the same action with different δ of 1, 3, 5, 10. Higher values for δ remove earlier trail of motion. 2nd row presents a running action. First 2 images are for δ=1 and the later pair for δ=3, while the 1st and 3rd images are taken mid-way and the 2nd and 4th images are taken at the end of the sequence. Fig. 4 shows the dependence on τ for an action having 26 frames.

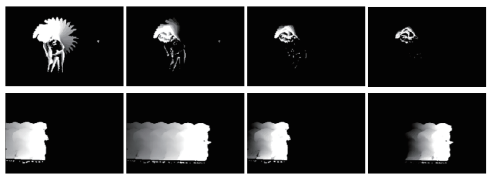

Figure 3. Dependence on δ in calculating MHIs.

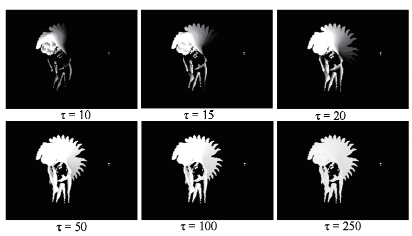

Figure 4. Dependence on τ to develop MHI images.

PR0002/09/0000-5401 ¥400 © 2009 SICE

For recognition, from MHI and MEI images, feature vectors are calculated by Hu moments [4]. However, Zernike moments, global geometric shape descriptors, Fourier transform, etc. can also be used for shape analysis from the templates. PCA can be used for dimension reduction of the feature vectors. For classification, SVM, *k*-nearest neighbor, Mahalanobis distance, Maximum Likelihood, etc. are employed.

However, it has some constraints: (i) its holistic generation (and matching) of the moment features computed from the *entire* template; (ii) occlusions of the body or errors from the implementation of the update function $\Psi(x,y,t)$; (iii) it is limited to only label-based recognition, where it could not yield any information other than specific identity matches; (iv) its requirement of having stationary objects, and the insufficiency of the representation to discriminate among similar motions; (v) its non-trajectory nature; (vi) it is unsuitable for dynamic background; (vii) it is not view-invariant; (vii) it cannot solve self-occlusion or overwrite problem.

3. APPROACHES BASED ON THE MHI

3.1 Various Approaches Employing the MHI Method

MHI and/or MEI templates are directly implemented without any significant modification by several researchers in different applications. Yau et al. [5] decomposed MHIs into wavelet sub-images in visual speech recognition. Moment-based features are extracted from sub-images to classify 3 consonants. A trajectory-guided recognition method is proposed [6], using Extended Kalman Filter and MHI. They tested the system to recognize various dynamic outdoor activities. T. Jan [7] collected trajectory information and its MHI for surveillance to decide possible suspicious behaviors. MHI is employed [8] for hand gesture recognition considering trajectories of the motion. [9] modeled action characteristics by MHI for hand gesture recognition in separating between events.

Gait Energy Image (GEI) [10] is developed like MEI for gait recognition. GEI is a gray-level image. Gait History Image (GHI) [11] for gait representation and recognition, inherits the idea of MHI. An adaptive camera navigation model is constructed for surveillance by automatically learning locations of high activity [12]. It measures activity from the MHI. Ahmad et al. [13] proposed silhouette energy image and silhouette history image to characterize motion for recognition. Edge motion history image (EMHI) [14] is computed by combining edge detection and MHI techniques from video sequence.

3.2. Variants of the MHI Method

Multiple-level MHI (MMHI) aims at overcoming the problem of motion self-occlusion by recording motion history at multiple time intervals [15]. It uses a simple bit-wise coding scheme and if motion occurs at time *t* at pixel location (x,y), it adds 2^{t-1} to the old motion value of MMHI. However, motion overwriting due to self-occlusion of MHI is robustly solved by Directional Motion History Image (DMHI) method [16]. In this approach, instead of background or frame subtraction, gradient-based optical flow is calculated between consecutive two frames and split it into four channels. Various complex actions and aerobics (which have more than one direction in its actions) are tested and achieved more than 92% recognition with DMHI, whereas, the MHI shows around 50% recognition result. The DMHI method requires four history templates and four energy templates for four directions. Various reduced feature vectors are proposed for DMHI to recognize motions fast with almost the same recognition result [17]. Also, DMHI performs well in low-resolution cases [18]. Moreover, a SVM-based system called Hierarchical Motion History Histogram (HMHH) is proposed [19] to solve motion overwriting problem.

Kellokumpu et al. [26] extract spatially enhanced local binary pattern (LBP) histograms from MHIs and MEIs, and model their temporal behavior with HMMs. They choose a fixed frame numbers in order to give a short term motion representation and the MHI is divided into four sub-regions through the centroid of the silhouette. All MHI and MEI LBP features are concatenated into one histogram and normalized so that the sum of the histogram equals one. This texture-based description of movements can handle overwrite problem of the MHI. One concern of this approach is the choice of the sub-regions division scheme for every action.

An advantage of MHI is that although it is a representation of the history of pixel-level changes, only one previous frame needs to be stored. However, at each pixel location, explicit information about its past is also lost in MHI when current change are updated to the model with their corresponding MHI values 'jumping' to the maximal value [23]. To overcome this problem, Ng and Gong [24] proposed a Pixel Signal Energy (PSE) in order to measure the mean magnitude of pixel-level temporal energy over a period of time defined by a backward window. The size of the backward window determines the number of frames (history) to be stored. Another recent development on MHI representation is Pixel Change History (PCH) [23]. This can measure the multi-scale temporal changes at each pixel. The PCH of a pixel ($P_{\varsigma,\delta}(x,y,t)$) is defined by,

$$P_{\varsigma,\delta}(x,y,t)=\begin{cases}\min\left(P_{\varsigma,\delta}(x,y,t-1)+\dfrac{255}{\delta},255\right) & \text{if } D(x,y,t)=1\\ \max\left(P_{\varsigma,\delta}(x,y,t-1)-\dfrac{255}{\delta},0\right) & \text{otherwise}\end{cases}$$

where, $D(x,y,t)$ is the binary foreground image, ς is an accumulation factor and δ is a decay parameter. When $D(x,y,t)$ =1, the value of a PCH increases gradually according to the accumulation factor, instead of jumping to the maximum value. In fact, MHI is a special case of PCH in that PCH image is equivalent to MHI when a parameter called accumulation factor (ς) is set to 1.

Davis [20] presented a new method for recognizing movement that relies on localized regions of motion derived from MHI. It gathers and matches multiple overlapping histograms of the motion orientations from MHI. It can handle variable-length movements as well as occlusion issue. Also, MHI is generalized by directly encoding the actual time in a floating point format, called *timed*-MHI [21]. Here, new silhouette values are copied in with a floating-point timestamp.

The MHI's limitations, relating to the 'global image' feature calculations can be overcome by computing dense local motion vector field directly from the MHI for describing the movement [22]. It extended basic MHI by transforming MHI into an image pyramid to permit efficient fixed-size gradient masks to be convolved at all levels of the pyramid to extract motion information at a wide range of speeds. The result is a hierarchy of motion fields, where resulting motion computed in each level is tuned to a particular speed (with faster speeds residing at higher levels). It is called hierarchical-MHI.

Temporal Motion Segmentation (TMS) method is proposed [16] that can segment a continuous action into its primitives. It can demonstrate the intermediate interpretation of complex motion into four directions, namely, right, left, up and down.

MHI can also be used to detect and interpret actions in compressed video data, by introducing motion flow history (MFH) that quantifies the motion in compressed video domain [25]. Motion vectors are extracted from the compressed MPEG stream by partial decoding. Then noise is reduced and the coarse MHI and the corresponding MFH are constructed at macro-block resolution instead at pixel resolution. In this way, they reduced the computation by 16 times. But using MFH, self occlusion or overlapping of motion on the image plane problem remains unsolved significantly.

3.3 View-invariant 3D Extensions of MHI Method

Several 3D extensions of basic MHI method are proposed for view-invariant 3D motion recognition [27-31]. Motion History Volume is introduced in 3D instead of the MHI for 2D. Albu et al. [27] presents a new 3D motion representation, called Volumetric-MHI (VMHI), to be used for the analysis of irregularities in human actions. The VMHI can be computed by,

$$VMHI(x,y,k)=\begin{cases} S(x,y,k)\nabla S(x,y,k+1) & \text{if } \aleph S(x,y,k)\neq\aleph S(x,y,k+1) \\ 1 & \text{otherwise} \end{cases}$$

where, $\aleph S(x,y,k)$ is the one pixel thick contour of the binary silhouette in frame k and ∇ stands for the symmetric difference operator. It tried to overcome limitations of MHI, related to motion self-occlusion, speed variability, and variable-length motion sequences. Shin et al. [28] presented a real-time gesture recognition model with 3D Motion History Model (MHM). Based on this work, Another view-invariant 3D recognition method, called Volume Motion Template (VMT) is proposed [31]. It extracts silhouette images using background subtraction and disparity maps. Then it calculates volume object in 3D space to construct VMT. With 10 gestures, they achieved good recognition result. Weinland et al. [29] described a 3D extension of MHI, called Motion History Volumes (MHV) based on visual hull for viewpoint-independent action recognition. In this approach, pixels are replaced with voxels, and voxel values in the MHV at time t are defined as:

$$v_\tau(x,y,z,t)=\begin{cases} \tau & \text{if } D(x,y,z,t)=1 \\ \max(0,v_\tau(x,y,z,t-1)-1) & \text{otherwise} \end{cases}$$

For feature extraction, 3D moments are employed [32].

4. MHI-BASED APPLICATION

The MHI method and the concept of the MHI and MEI templates are widely employed and analyzed. Table I at the end of the paper summarizes various applications based on the MHI method into three broader groups, broadly: (i) gesture or action recognition; (ii) motion analysis, and (iii) interactive systems. This Table contains the employed databases or actions sets, experimental results, important features, classification schemes, etc. The MHI and MEI present region of interest in a moving scene in two different formats and these cues are exploited in various applications.

From this Table, it is evident that MHI is widely employed by many researchers in computer vision, due to the fact that the creation of MHI and hence the segmentation of motion region from the scene can be computed in real-time and easily. Some applications are achieved based on developments of the MHI method.

5. DISCUSSIONS

We pointed some constraints in Section 2 and Subsections 3.2 and 3.3 added variants and developments of MHI method to solve some of these limitations. Motion overwriting due to motion self-occlusion problem is addressed by some researchers. The DMHI method solved this problem significantly. However, MHI's problems related to multiple moving subjects in the scene, moving background or moving camera, etc. are still unaddressed properly. When two or more than two persons are in-view, these methods can not recognize properly, especially when all of them are moving in different directions.

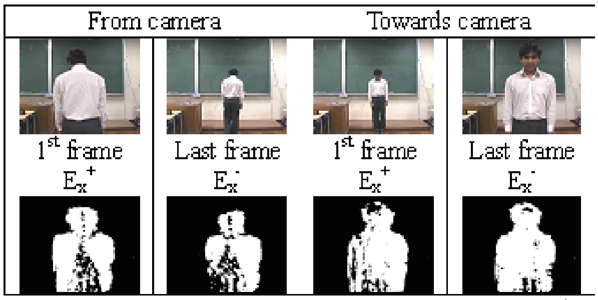

Figure 5. Scenarios of walking from camera (1st 2 columns), & towards camera (last 2 columns).

Moreover, MHI and its variants can not recognize if the person is walking towards the camera's optical axis or if it moves something like diagonal directions. Fig. 5 shows the case for two actions: (i) a person is walking from the camera to move outward-direction with the optical axis line; and (ii) a person is walking towards the camera from far, both almost in-line with the optical axis of the camera. It is evident from the energy images that the system can not profoundly separate these two actions. So this issue should be solved with some semantics.

Another important pair of activities are running and walking across the optical axis. And to recognize these or distinguish walking motion from running motion for video surveillance is very difficult with the present format of the MHI method. Refs. [10-11,55] proposed GEI, GHI and AEI respectively for gait analysis [10-11] and recognizing walking and running motion. However, Ref. [55] has limited the actions datasets to prove it robustly. One intuitive way to achieve better features for separating walking and running motion is to employ the DMHI method and use the decay parameter (δ) as a higher value so that we can achieve the ripples at the top of the template and use these features for recognition.

As stated above, MHI's problem related to the presence of multiple moving persons/objects in the scene is still unsolved, though it is a real concern for multiple object identification [56]. Image depth analysis can aid to solve this problem. Researchers may think about the camera movement and its effect. Usually, camera motion compensation is difficult and the effect of camera movement and the employment of MHI are not solved, though Davis et al. [12] applied MHI with PTZ camera. Instead of 2D view-based approaches, some 3D view-invariants approaches [27-31] are proposed. But these are highly constrained by the use of multiple cameras, extra pre-processing loads and computational cost.

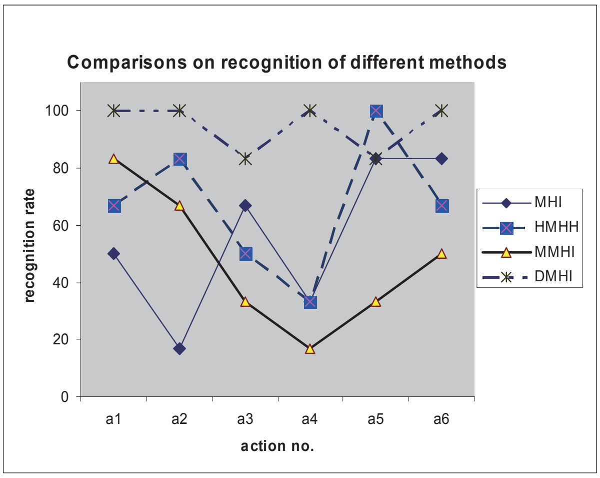

Figure 6. Comparative results for different methods.

The MHI [3], MMHI [15], HMHH [19] and DMHI [16] methods are experimentally compared to recognize some exercise activities that constitute motion overlapping and self-occlusion [54]. In this comparative analysis, the DMHI method outperforms other methods to solve self-occlusion problem. Fig. 6 shows the comparative recognition results for these methods [54].

6. CONCLUSIONS

This paper presented the variants of motion history image (MHI) method, its application realms by various researchers and some analysis. It is well-known that action analysis, recognition, tracking, etc. are very challenging tasks, due to large variations in human motion and appearance, camera viewpoint and environment settings, etc. [57]. The field of action and activity representation and recognition is relatively old, yet still immature [2]. Some important but common motion recognition problems are still unsolved properly by the computer vision community and hence action and activity analysis and recognition approaches are still in its infancy. For MHI approaches, proper moving area segmentation from the scene is important and hence the employment of the update function in calculating the MHI or others is important. Though for simple activities, background subtraction or frame differencing approaches are considered, it is important to ponder about these according to the environment. Cases related to dynamic background or moving camera creates difficulties and therefore, deployment of the MHI templates in these dynamic cases are part of tasks for future researchers to solve. We think that technologies with multiple clues (e.g., motion, shape, edge information [14], color or texture, gradient information [58], motion vectors, etc.), or a fusion of information will produce better result. Moreover, for recognition purposes, proper usage of classification approach is another important key. We hope that this paper will be beneficial to various researchers to understand the MHI method, its variants and applications areas.

Reference

1. M.A.R. Ahad, J.K. Tan, H. Kim, and S. Ishikawa, "Human activity recognition: various paradigms", Proc. ICCAS, 2008, pp. 1896-1901.
2. T. Moeslund, A. Hilton, V. Kruger, "A survey of advances in vision-based human motion capture and analysis", CVIU, 104, 2006, pp. 90-126.
3. A. Bobick, J. Davis, "The recognition of human movement using temporal templates", IEEE Trans. PAMI, 23:3, 2001, pp. 257-267.
4. M. Hu, "Visual pattern recognition by moment invariants", IRE Trans. on Info. Theory, 1962, 179-187.
5. W. Yau, D. Kumar, S. Arjunan, S. Kumar, "Visual speech recognition using image moments and multiresolution wavelet", Conf. on Computer Graphics, Imaging and Visualization, 2006, pp. 194-199.
6. R. Rosales, S. Sclaroff, "3D trajectory recovery for tracking multiple objects and trajectory guided recognition of actions", CVPR, 2, 1999, pp. 117-123.
7. T. Jan, "Neural network based threat assessment for automated visual surveillance", IEEE Joint Conf. on Neural Networks, 2, 2004, pp. 1309-1312.
8. C. Shan, Y. Wei, X. Qiu, T. Tan, "Gesture recognition using temporal template based trajectories", ICPR, 3, 2004, pp. 954-957.
9. K. Alahari, C. Jawahar, "Discriminative actions for recognizing events", ICVGIP, 2006, pp.552-563.
10. J. Han, B. Bhanu, "Individual recognition using gait energy image", IEEE Trans. PAMI, 28:2, 2006, pp. 316-322.
11. J. Liu, N. Zhang, "Gait history image: a novel temporal template for gait recognition", IEEE ICME, 2007, 663-666.
12. J. Davis, A. Morison, D. Woods, "Building adaptive camera models for video surveillance", WACV, 2007.
13. M. Ahmad, M. Hossain, "SEI and SHI representations for human movement recognition", ICCIT, 2008, pp. 521-526.
14. D. Chen, J. Yang, "Exploiting high dimensional video features using layered Gaussian mixture models", IEEE ICPR, 2006, 4 pages.
15. M. Valstar, M. Pantic, I. Patras, "Motion history for facial action detection in video", Conf. SMC, 2004, pp. 635-640.

16. M. Ahad, J. Tan, H. Kim, S. Ishikawa, “Temporal motion recognition and segmentation approach”, Int. J. of Imaging Systems & Technology, 19, 2009, pp. 91-99.
17. M. Ahad, T. Ogata, J. Tan, H. Kim, S. Ishikawa, “Template-based human motion recognition for complex activities”, IEEE Conf. on SMC, 2008, pp. 673-678.
18. M. Ahad, J. Tan, H. Kim, S. Ishikawa, “Directional motion history templates for low resolution motion recognition”, Proc. IEEE IECON, 2008, pp. 1875-1880.
19. H. Meng, N. Pears, C. Bailey, “A human action recognition system for embedded computer vision application”, Workshop on Embedded Computer Vision (with CVPR), 2007, pp. 1-6.
20. J. Davis, “Recognizing movement using motion histograms”, MIT Media Lab. Tech. Report No. 487, 1998.
21. G. Bradski, J. Davis, “Motion segmentation and pose recognition with motion history gradients”, Machine Vision and Applications, 13:3, 2002, pp. 174-184.
22. J. Davis, “Hierarchical motion history images for recognizing human motion”, IEEE Workshop on Detection and Recognition of Events in Video, 2001, pp. 39-46.
23. T. Xiang, S. Gong, “Beyond tracking: modelling activity and understanding behaviour”, IJCV, 67:1, 2006, pp. 21-51.
24. J. Ng, S. Gong, “Learning pixel-wise signal energy for understanding semantics”, Proc. BMVC, 2001, pp. 695-704.
25. R. Babu, K. Ramakrishnan, “Compressed domain human motion recognition using motion history information”, Proc. ICIP, 2, 2003, pp. 321-324.
26. V. Kellokumpu, G. Zhao, M. Pietikäinen, “Texture based description of movements for activity analysis”, Proc. Conf. Computer Vision Theory & Application, 2008, pp. 368-374.
27. A.B. Albu, T. Beugeling, “A three-dimensional spatiotemporal template for interactive human motion analysis, J. of Multimedia, 2:4, 2007, pp. 45-54.
28. H. Shin, S. Lee, S. Lee, “Real-time gesture recognition using 3D motion history model”, Proc. Conf. on Intelligent Computing, Part I, LNCS 3644, 2005, pp. 888-898.
29. D. Weinland, R. Ronfard, E. Boyer, “Free viewpoint action recognition using motion history volumes”, Computer Vision Image Understanding, 104:2, 2006, pp. 249-257.
30. C. Canton-Ferrer, J. Casas, M. Pardas, “Human model and motion based 3D action recognition in multiple view scenarios”, Conf. European Signal Process., 2006, pp. 1-5.
31. M. Roh, H. Shin, S. Lee, S. Lee, “Volume motion template for view-invariant gesture recognition”, Proc. ICPR, 2, 2006, pp. 1229-1232.
32. C. Lo, H. Don, “3-D moment forms: their construction and application to object identification and positioning”, IEEE Trans. PAMI, 11:10, 1989, pp. 1053-1063.
33. S. Kumar, D. Kumar, A. Sharma, N. McLachlan, Classification of hand movements using motion templates and geometrical based moments, in: Proc. Int’l Conf. on Intelligent Sensing and Information Processing, 2003, pp. 299-304.
34. J. Davis, Sequential reliable-inference for rapid detection of human actions, in: Proc. IEEE Workshop on Detection and Recognition of Events in Video, July 2004, pp. 1-9.
35. K. Leman, G. Ankit, T. Tan, PDA-based human motion recognition system, Int. J. Software Engineering and Knowledge, 2:15(2005) 199-205.
36. R. Singh, B. Seth, U. Desai, “A real-time framework for vision based human robot interaction”, IEEE/RSJ Conf. on Intelligent Robots and Systems, 2006, pp. 5831-5836.
37. T. Ogata, J. Tan, S. Ishikawa, “High-speed human motion recognition based on a motion history image and an Eigenspace”, IEICE Trans. IS, E89-D:1, 2006, pp. 281-289.
38. W. Ryu, D. Kim, H.-S. Lee, J. Sung, D. Kim, “Gesture recognition using temporal templates”, ICPR, Demo, 2006.
39. J.K. Tan, S. Ishikawa, “High accuracy and real-time recognition of human activities”, Annual Conf. of IEEE Industrial Electronics Society, 2007, pp. 2377-2382.
40. S. Vitaladevuni, V. Kellokumpu, L. Davis, “Action recognition using ballistic dynamics”, CVPR, 8 pages, 2008.
41. C. Orrite, F. Martınez, E. Herrero, H. Ragheb, S. Velastin, “Independent viewpoint silhouette-based human action modelling and recognition”, Proc. MLVMA, 2008, pp. 1-12.
42. J. Piater, J. Crowley, “Multi-modal tracking of interacting targets using Gaussian approximations”, Workshop on PETS with CVPR, 2001, pp. 141-147.
43. J. Ruiz-del-Solar, P. Vallejos, “Motion detection and tracking for an AIBO robot using camera motion compensation and Kalman filtering”, Proc. RoboCup Int. Symposium 2004, LNCS 3276, 2005, pp. 619-627.
44. D. Son, T. Dinh, V. Nam, T. Hanh, H. Lam, “Detection and localization of road area in traffic video sequences using motion information and fuzzy-shadowed sets”, Proc. IEEE Int. Symposium Multimedia, 2005, pp. 725-732.
45. Z. Yin, R. Collins, “Moving object localization in thermal imagery by forward-backward MHI”, Proc. IEEE Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum, 2006, pp. 133-140.
46. L. Li, Q. Zeng, Y. Jiang, H. Xia, “Spatio-temporal motion segmentation and tracking under realistic condition”, Proc. Int’l Symposium on Systems and Control in Aerospace and Astronautics, 2006, pp. 229-232.
47. I. Petrás, C. Beleznai, Y. Dedeoğlu, M. Pardàs, et al., “Flexible test-bed for unusual behavior detection”, Proc. ACM Conf. Image and Video Retrieval, 2007, pp. 105-108.
48. J. Davis, A. Bobick, “Virtual PAT: a virtual personal aerobics trainer”, Perceptual User Interface, 1998, pp. 13-18.
49. J. Davis, G. Bradski, “Real-time motion template gradients using Intel CVLib”, Proc. ICCV Workshop on Frame-rate Vision, 1999, pp. 1-20.
50. A. Bobick, S. Intille, J. Davis, et al., “The Kidsroom: a perceptually-based interactive and immersive story environment”, Presence: Teleoperators and Virtual Environments, 8:4, 1999, pp. 367-391.
51. M. Pantic, I. Patras, M. Valstar, “Learning spatio-temporal models of facial expressions”, Proc. ICMB, 2005, pp. 7-10.
52. Q. Nguyen, S. Novakowski, J. Boyd, C. Jacob, G. Hushlak, “Motion swarms: video interaction for art in complex environments”, Proc. ACM Int. Conf. Multimedia, 2006, pp. 461-469.
53. A. Senior, S. Tosunoglu, “Hybrid machine vision control”, Proc. Florida Conf. on Recent Advances in Robotics, 2005, pp. 1-6.
54. Md. Atiqur Rahman Ahad, J.K. Tan, H. Kim, S. Ishikawa, Analysis of motion self-occlusion problem due to motion overwriting for human activity recognition, Journal of Multimedia, 4, 2009, 11 pages.
55. V. Chandrashekhar, K. Venkatesh, “Action energy images for reliable human action recognition”, Proc. Asian Symposium on Information Display, 2006, pp. 484-487.
56. C. Cedras, M. Shah, “A survey of motion analysis from moving light displays”, Proc. CVPR, 1994, pp. 214-221.
57. R. Poppe, “Vision-based human motion analysis: an overview”, Computer Vision and Image Understanding, 108:1-2, 2007, pp. 4-18.
58. S. McKenna, S. Jabri, Z. Duric, H. Wechsler, A. Rosenfeld, “Tracking groups of people”, Computer Vision and Image Understanding, 80:1, 2000, pp. 42-56.

Table 1. Various Applications by employing MHI and its variants.

[Ref.]	Employed datasets/applications	Results/comments/features
(i) Applications related to various motion/action recognition		
[6]	8 actions by 7 persons	PCA, Hu moment, KNN, Gaussian & Gaussian Mixture classifier
[20]	Recognition in real-time	Histograms of motion orientations of the MHI
[3]	18 aerobic exercises by 1 instructor – taken several times	Good rec. rate; Mahalanobis distance; Hu moments
[25]	7 actions by 5 subjects	Good rec. rate; Compressed domain motion; KNN, MLP
[33]	5 hand movements for 10 repetitions, by 5 persons	Hu moments; rec. rate is 96%; classifier: back propagation based multilayer perception ANN
[34]	Walk, run, stand, by 3 persons, 8 viewpoints, thermal camera	Recognition rate is 77%
[8]	Real-time hand gesture recognition	Tracking by Mean Shift Embedded PF
[28]	Real-time gesture recognition	View-invariant, 3D approach
[35]	PDA-based recognition system	Employ PDA and limited scope
[36]	11 gestures for robot behavior in real-time	Motion Color Cue with MHI, MEI
[31]	10 gestures	Good rec. rate; 3D
[37]	6 actions by 9 persons – taken 3 times	Eigenspace, SMI
[30]	5 calibrated wide-lens cameras, in SmartRoom	Real-time recognition, Bayesian classifier
[29]	IXMAS dataset, 11 actions	View-invariant action recognition, 3D approach
[14]	6 different shots of 1 hr video from CNN	PCA, LGMM, SVM
[38]	10 different gestures	Recognition rate is 90%, indoor environment, using a stereo camera for a robot
[9]	Marcel's Dynamic Hand Gesture database is employed for hand gesture recognition, and 4 human actions	32% average reduction in error is achieved for some event pairs.
[10-11]	USF HumanID gait DB; CMU, UMD, CASIA gait DB	Good comparative results
[39]	6 actions from 4 cameras by 11 subjects	Poor result: as motion overwriting in some actions
[19]	6 actions by 25 persons	Good rec. rate; SVM
[16]	10 aerobics by 8 persons; 5 actions by 9 persons	Good rec. rate.; Optical flow, KNN, Hu Moment
[40]	14 gestures by 5 persons; 7 video sequences by 6 subjects	Good rec. rate; Fourier-based MHI, DP
[26]	15 gestures by 5 persons; 10 actions by 9 persons	Good result; LBP histogram, HMM
[13]	Korea University gesture DB, part of KTH DB	Hu and Zernike moments, global geometric shape descriptors, PCA, SVM, KNN, Bayes
[41]	Virtual Human Action Silhouette (ViHASi) DB, IXMAS DB	SOM, Ml classifier, HMM, 77% rec. rate (IXMAS)
(ii) Applications related to motion analysis, tracking, etc.		
[6]	Dynamic outdoor activities	Trajectory-guided tracking by Extended Kalman Filter
[42]	Tracking, used first three sequences of PETS dataset	Concerns on Gaussian weighting, Kalman filter
[43]	Motion detection and tracking for an AIBO robot	Camera motion compensation, Kalman filtering
[7]	Threat assessment for surveillance in car parking	Tracking, NN-based, not promising result
[53]	Tracking with CAMSHIFT algorithm	Neural network is employed
[44]	Detection and localization of road area in traffic video	Fuzzy-shadowed sets are used
[23]	Activity analysis and recognition in indoor and outdoor scenes	EM, BIC, DPN, Dynamically Multi-Linked-HMM
[45]	Moving object localization from thermal imagery	Used RANSAC
[46]	Moving object tracking	Used MHI though overall method is not great
[12]	Adaptive camera models for video surveillance using PTZ camera, outdoor in various places	Need to include more features (e.g., texture, color, etc.) for improvement
[27]	Analysis of sway and speed-related abnormalities of human actions; 5 different actions	Solves motion self-occlusion, action length variability,
[47]	Real-time detector for unusual behavior, 4 major partners (ACV, BILKENT, UPC and SZTAKI) achieved this task	Employed 3D-MHI of [30], web server-based real-life detection module, tracking, outdoor
[16]	Temporal motion segmentation and action analysis	Indoor and outdoor experiment, need to implement in an intelligent robot
(iii) Development of various interactive systems based on the MHI method		
[48]	Virtual aerobics trainer	It watches and responds to the user as he/she performs the workout
[20]	Interactive art demonstration	Map different colors to the various timestamps within the MHI for fun
[50]	The KidsRoom	An interactive, narrative play space for children, with virtual monsters
[15,51]	21 or less facial Action Unit classes; MMI-Face-DB; Cohn-Kanade Face DB	Poor recognition rate; preprocessing loads
[52]	Interactive art demonstration	In complex environment
[5]	Speech recognition – 3 consonants	SWT; very limited result