International Journal of Innovative
Computing, Information and Control
Volume **8**, Number **2**, February **2012**

 ISSN 1349-4198
pp. 1359–1370

AN EFFECTIVE DIRECTIONAL MOTION DATABASE ORGANIZATION FOR HUMAN MOTION RECOGNITION

S. M. Ashik Eftakhar, Joo Kooi Tan, Hyoungseop Kim and Seiji Ishikawa

Department of Mechanical and Control Engineering
Kyushu Institute of Technology
Sensui-cho 1-1, Tobata, Kitakyushu, Fukuoka 804-8550, Japan
{ ashik; etheltan; ishikawa }@ss10.cntl.kyutech.ac.jp

Received November 2010; revised March 2011

Abstract. *Automatic recognition of human motions has an increasing demand in the recent visionary world. However, with the registration of large number of motions from varying viewpoints, the necessity for an effective motion database for recognition has become a vital issue. In the context of motion database development, this paper proposes a directional database organization for human motion recognition. This organization partitions the motion database into several sub-databases on the basis of camera orientation. Separate feature spaces are constructed, and correspondingly directional sub-databases are built, leading to the constitution of the complete motion database. The directionally similar but semantically different motions are properly distinguished. To show the robustness of the proposed organization for recognizing human motions, a set of motions captured from varying viewpoints is analyzed. An eigenspace representation is employed as a generic feature space that sufficiently characterizes the motion features. Motion History Image (MHI) and Exclusive-OR (XOR) image representations are used as motion templates where MHI is found performing better than XOR image. The experimental results show high-level of satisfactory performance and claim the significant improvement over earlier developed systems.*
Keywords: Directional database organization, Eigenspace, Motion history image, Exclusive-OR image, Camera orientation

1. **Introduction.** In recent years, with the increasing interest in the field of computer vision and image processing, the development of an efficient human motion recognition system has become an indispensible part of the intelligent systems and Human-Computer Interaction (HCI) systems. Developing a reliable intelligent system that is capable of manifesting what a human is performing in a scene is very much challenging task. This sort of system has wide variety of applications, especially in surveillance, virtual and augmented reality, animation, intelligent robots, diagnostics of orthopedic patients in clinics and hospitals, supporting aged people in rehabilitation centers, performance evaluation and training of the athletes in sports, and so on. Due to diverse applications for such systems, it requires robustness as well as accuracy. Moreover, the system is subjected to be in use in real-time; this urges for relatively fast response of the system. This attribute constrains on the system development that the system should have the capability of high-speed recognition. Therefore, having a number of aspects for the recognition system, the literature related to the problem of recovering and recognizing human motions in a scene is intensive [1-7]. However, we focus on the methods addressing the specific problem of recognizing human motions from image sequences without using markers, tracking devices, or special body suits. Based on whether or not a priori knowledge about the object's shape is required, the methods for human motion analysis can be classified broadly

into two categories: model-based and appearance-based approaches [10,11]. However, other forms of categories are also available [12]. Both the approaches have their own advantages and disadvantages. Appearance-based approaches are applicable to diverse situations, since they do not require a specific object model. Those methods are sensitive to noise in general, because they lack any mechanism to distinguish noise from signal in visual input. Appearance-based approaches build a body representation in a bottom-up fashion by first detecting appropriate features in an image, whereas model-based approaches build the body representation by fitting to the image data using some predefined parameter values of the parametric body model [8,9]. In practice, motion-specific representation composed of adequate features to represent each motion uniquely is enough to accomplish the task of recognition. Therefore, advanced image processing techniques are being comprehensively investigated in search of effective representation of a motion. Standard techniques for the motion representation include the ones based on Motion History Image (MHI) and its variants [4,11,13,14]. Motion history-based representations include not only the movement of a body itself but also the change of position of a person in a scene. However, object's silhouette information alone can be used as an input for a recognition system. Wang and Suter [5] used silhouettes as the input to their recognition system. Elgammal and Lee [6] also used silhouettes without motion history. Moreover, another motion recognition approach was also proposed which considered multi-view motion representation and recognition [18]. In this approach, the motion postures are iteratively transformed into a single eXclusive-OR (XOR) template image for the task of registration and recognition.

However, in order to deal with the high dimensional complex information extracted from human motion, it is necessary to find reduced representation of the motion while maintaining sufficient discriminating data for performing the recognition. To accomplish these goals, existing researches have used simple data reduction techniques such as Principal Component Analysis (PCA) [4], Eigenspace technique [19] and Locality Preserving Projections (LPP) [5,15]. Moreover, a statistical matching method using Hidden Markov Model (HMM) that allows for a principled probabilistic modeling of the temporal sequential information is also adopted in various works [6,22]. An alternative approach for matching the data sequences using Dynamic Time Warping (DTW) is also employed in recent works [2,21]. The recognition methods mentioned in the aforesaid literature commonly use single or frontal cameras to capture motions in the case of view-based motion analysis. However, an intelligent system may have the flexibility of orientation-independent recognition. Therefore, it is necessary to handle the orientation-specific data in an effective way.

In this paper, we propose a human motion recognition approach capable of distinguishing the orientation-specific motions effectively by means of a structured data organization. The novelty of the proposed approach lies in improving the precision and robustness of the recognition by making use of the directional organization of the motion database (referred to as "directional motion sub-database") corresponding to the varying viewpoints and the nearest index searching strategy with the database. The directional candidates obtained from the directional motion sub-databases play an effective role to find the similar motions. Unlike earlier researches, we have proposed an adoption of multi-viewpoint concept without integrating the orientation-wise information, and thus reducing the load of detailed analysis.

2. Motion Representation.

View-based motion representation is the demonstration of change in brightness values within the consecutive motion frames. Due to the diversity in applications, different motion representation approaches have been adopted since many years. Among those, the representation adopted in [14], known as Motion History Image, is an effective motion representation that can trace the spatial and temporal information within each motion. Besides, the exclusive-OR representation proposed in [18] performs well in several motion recognition applications. We shall describe our proposed system using these two well-recognized representation techniques and perform experiments to show the efficiency and compatibility between these two techniques.

2.1. **Motion history image (MHI).** Motion History Image is a view-specific representation of movement, where movement is defined as motion over time. MHI, as the name implies, keeps track of the motion history, i.e., representing how the motion is moving along a certain period of time. This is very famous and well-established motion representation strategy since many years. This is represented as a frame-based temporal template for human motion. In generating MHI, the temporal information is specified by the pixel intensity. Let $H_\tau(x, y, t)$ be the pixel intensity function of the temporal history of motion at a particular point (x, y) on an image. The function is represented as shown in (1).

$$H_\tau(x,y,t) = \begin{cases} \tau & \text{if } D(x,y,t) = 1 \\ \max(0, H(x,y,t-1) - 1) & \text{Otherwise} \end{cases} \tag{1}$$

Here, $D(x, y, t)$ is a difference image in a binarized form constructed from successive frame difference. The function $H_\tau(x, y, t)$ returns a scalar value. According to the function, the more recently moving pixels are brighter than the past moving pixels in the generated MHI. In (1), τ is taken as the temporal extent which is critical to define. But for the flexibility of the value of τ, it can be taken as the maximum gray level pixel value (255) or the maximum number of frames defining a motion. An example of a MHI is shown in Figure 1.

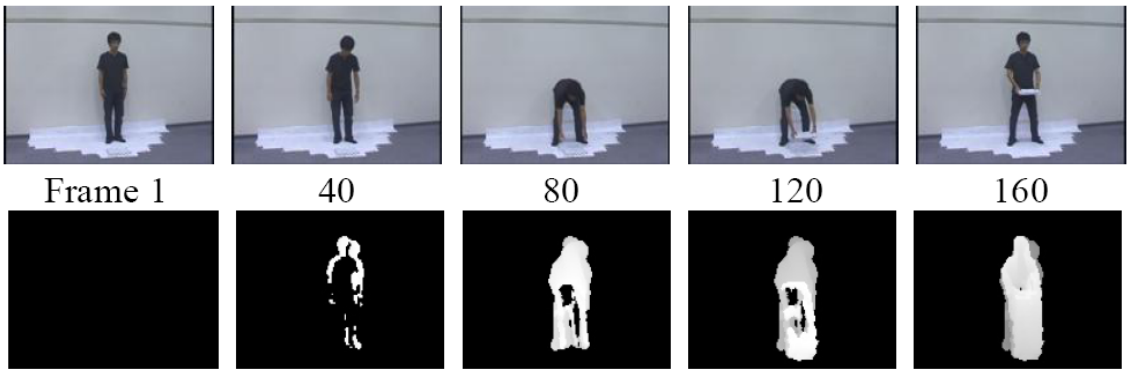


FIGURE 1. *Carry-up* motion and a corresponding MHI: Top row contains key frames and the bottom row is the process of creating a MHI starting from frame 1

2.2. **Exclusive-OR (XOR) image.** The exclusive-OR representation is generated by simple logical operation within successive motion frames. The concept of this representation was originally proposed by J. K. Tan [18] where the compressed XOR motion images formed an effective motion database. It is a cumulative exclusive-OR form of the motion frame representing the motion features by a single motion template. Each motion frame is binarized by using a fixed thresholding (e.g., threshold value = 20) function. Referring to (2), it is applied to every binarized frame within the motion frame set for constructing the XOR motion template.

$$\begin{aligned} &U_c^{m,h}(2) = f_c^{m,h}(1)\ XOR\ f_c^{m,h}(2) \\ &U_c^{m,h}(r) = U_c^{m,h}(r-1)\ XOR\ f_c^{m,h}(r) \quad r = 3, 4, \cdots, R \\ &U_c^{m,h} \equiv U_c^{m,h}(R) \end{aligned} \tag{2}$$

In (2), the XOR image of motion m of person h obtained from camera c is denoted by $U_c^{m,h}(r)$, whereas f, U denote *binarized motion frame* and *XOR frame*, respectively. Hence, for M distinct motions of H persons each motion having Rframes from Ccamera directions generate a motion database of MHC XOR images. This is clearly an appearance-based motion representation method representing one template image corresponding to a single motion. This method is capable of eliminating the effect of background or stationary things in a scene. Only the moving portion in the scene is taken into account by using the logical XOR operation. An example of the generation of XOR representation is illustrated in Figure 2.

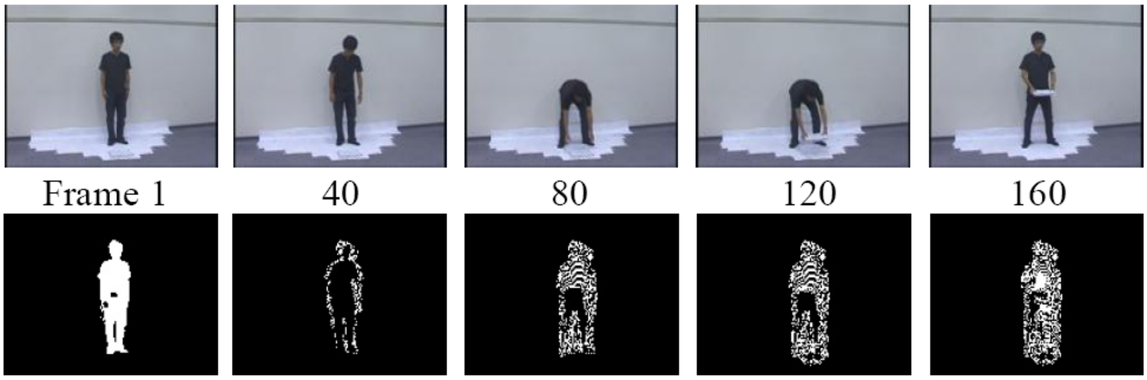


FIGURE 2. *Carry-up* motion and a corresponding XOR image: Top row contains key frames and the bottom row is the process of making an Exclusive-OR image starting from frame 1

3. **Construction of Directional Feature Spaces.**

A feature is a salient attribute for characterizing a motion and determining the similarity of it among a number of training motions [19]. The eigenvectors corresponding to the prominent eigenvalues serve this purpose by constructing an eigenspace of projected motion data. An eigenspace is a high-dimensional feature space that represents the proximity among the set of data. It is a modified form of Karhunen-Loeve Transform (KLT) that is used to derive relationship among different random variables. In practice, a large set of learning motions is required to be projected onto the eigenspace by finding prominent eigenvectors. For each motion image (MHI or XOR image) $I_m\,(m = 1, 2, \cdots, M)$, an image matrix is defined and the brightness is also normalized to minimize its effect. A data matrix $\boldsymbol{X}$ is obtained by subtracting the average image $\boldsymbol{c}$ from each motion image set as follows,

$$\boldsymbol{X} = (\boldsymbol{x}_1 - \boldsymbol{c}, \boldsymbol{x}_2 - \boldsymbol{c}, \cdots, \boldsymbol{x}_M - \boldsymbol{c}) \tag{3}$$

where $\boldsymbol{c} = \dfrac{1}{M} \sum\limits_{m=1}^{M} \boldsymbol{x}_m$.

The image matrix $\boldsymbol{X}$ has the order $N \times M$, where M is the total number of motion images, and N is the total number of pixels in each image. To compute eigenvectors of the motion image set, a covariance matrix $\boldsymbol{Q}$ is defined as:

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \tag{4}$$

The eigenvalues and corresponding eigenvectors are calculated from the $N \times N$ covariance matrix $\boldsymbol{Q}$ by solving the eigenvalue problem. Among N eigenvectors, k most prominent eigenvectors are chosen using *Principal Component Analysis* (PCA) to create an eigenspace *ES* consisting of the training motions [20]. For each camera viewpoint, separate eigenspaces are created, which we call directional eigenspaces, by projecting corresponding motions onto those using (5). Equal number of sub-databases are also built and maintained. Each sub-database returns a single *candidate* motion for a motion query. Moreover, a global eigenspace containing all the learning motions is also built and maintained to decide the most similar one among several candidate motions. The global eigenspace is constructed in a similar fashion as directional eigenspaces.

$$\boldsymbol{g}_m = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_k)^{\mathrm{T}}(\boldsymbol{x}_m - \boldsymbol{c}) \tag{5}$$

4. Development of a Motion Database.

A database, solely, relies upon the organization of data within the computer memory. The most common database organization is linear, i.e., the data arranged in the order of its input. At the time of query, the database performs sequential blind searching among the data. In order to overcome the problem of sequentiality in the query, many researchers have been comprehensively involved in the development of a suitable database that is organized in a non-sequential manner and also capable of quick and successful retrieval. Moreover, due to the increased number of motion archives, the maintenance of the database organization is also drawing much attention. As a result, the B-tree [17] database structure is adopted in our research as a structured non-sequential motion database.

4.1. Construction of structured database.

In order to develop a structured database, the motions which are projected onto the eigenspaces are required to be indexed into a numeric format for the flexibility of storage. The dimension of an eigenspace is taken as an important cue in space partitioning. The eigenspace is uniformly divided into several divisions. Each eigen-axis $\boldsymbol{e}_k$ $(k = 1, 2, \cdots, K)$ is partitioned into S $(S > 1; \mathit{integer})$ divisions each having equal edge length of L. Each hypercube with edge length L along each eigen-axis is referred to as *a bin* in a physical form, and *an index* in a numeric form. In this paper, the edge length L of the hypercube is termed as *bin length*. Each motion point, which we call bin or index, is assigned a digit from 0 to S-1 along each eigen-axis [20]. Therefore, an index becomes a K-digit S-nary number. However, in the case of no division of the space, in fact, no database system exists; it is rather sequential storage of the motion data. Therefore, the structured form of the database is preferable to realize a profound database system.

4.2. Directional organization of structured database.

In the case of the motions with several orientations, the motions can be grouped into several motions sets based on orientation. The steps for constructing the directionally organized structured database are as follows:

a. Capture the training motions having c $(c > 1; \mathit{integer})$ orientations by maintaining motion synchronization.
b. Create motion sets M_i $(i = 1, 2, \ldots, c)$ based on the orientation.
c. Construct eigenspaces ES_i corresponding to each motion set M_i using the scheme described in Section 3.
d. Construct the structured B-tree sub-database BSB_i corresponding to each eigenspace ES_i taking the division parameter S as described in Section 4.1.
e. Combine all the sub-databases to develop directionally organized database.

4.3. **Robustness of the directional organization.** The B-tree structured database maintains the ordered arrangement of data within the tree structure [16-18]. Each index, as generated in Section 4.1, is also assigned a decimal value based on the radix of the index. Depending upon the assigned value, the indexes are stored in an orderly way within the database. The B-tree retrieval algorithm is then applied to retrieve the matched index, or the most appropriate position if matching fails. However, the decimal value does not efficiently represent the neighboring indexes of an index corresponding to its co-ordinate values within original feature space. Moreover, till now there is no standard algorithm for the indexing strategy to select the nearest neighboring point within the space when the exact match of an index is not encountered. Our adopted approach is to calculate the digit-wise *Sum of Squared Difference* (SSD) between the consecutively stored indexes within the B-tree. However, it is not the exact measure, rather an approximation, to select the nearest index within the space. The nearest index searching algorithm is described below. In the algorithm, *Keys*(..) denote the pointer to the successor (or children) within the tree structure.

Step-1. For query index $\boldsymbol{y}$, $\boldsymbol{y > x_i}$ and $\boldsymbol{p_{i-1}}$=NULL
- Calculate $MinDist(MinDist(x_{i-1}, y),\ MinDist(x_i, y))$

Step-2. For query index $\boldsymbol{y}$, $\boldsymbol{x_i > y \geq x_i}$ $\boldsymbol{p_i}$=NULL
- Calculate $MinDist(MinDist(x_i, y),\ MinDist(x_{i+1}, y))$

Step-3. For query index $\boldsymbol{y}$, $\boldsymbol{y < x_i}$ and $\boldsymbol{p_{i-1} \neq}$ NULL
- Calculate $MinDist(MinDist(x_{i-1}, y),\ MinDist(x_i, y))$ as MD_1
- Calculate $\underset{l=1,2,3,\cdots}{MinDist}(x_{p_{i-1}^{(l)}}, y)$ as MD_2
- Calculate $MinDist(\mathrm{MD}_1, \mathrm{MD}_2)$

Step-4. For query index $\boldsymbol{y}$, $\boldsymbol{x_i > y \geq x_i}$ $\boldsymbol{p_i \neq}$ NULL
- Calculate $MinDist(MinDist(x_{i+1}, y),\ MinDist(x_i, y))$ as MD_1
- Calculate $\underset{l=1,2,3,\cdots}{MinDist}(x_{p_i^{(l)}}, y)$ as MD_2
- Calculate $MinDist(\mathrm{MD}_1, \mathrm{MD}_2)$

The conventional database organization is represented by high dimensional eigenspace, whereas the directional organization is represented by comparatively low dimensional space due to the splitting up the whole dataset into directionally independent datasets. It is certain that with the increase in number of dimensions, the probability of miss-selection proportionally increases. Thus, the conventional organization exhibits lower possibility to select the exact nearest index than that of the directional organization based on the aforementioned searching algorithm. Therefore, our proposed directional organization theoretically and experimentally proves the robustness of the recognition system.

5. **Recognition.** The recognition strategy is quite simple, but effective. When an unknown motion comes, it is first represented as a sequence of image frames and is processed for generating motion representation. A MHI or XOR image is generated from the motion frames, and projected onto each directional eigenspace. An index, representing motion identity within the directional sub-database, is generated from the point corresponding to the unknown motion after its projection onto each eigenspace. For each camera orientation, the equal number of similar motions is obtained by searching the corresponding B-Tree sub-database as mentioned in Section 4.3. Thus we get several candidate motions for an unknown motion. These candidate motions are projected onto the global eigenspace as $\mathbf{g}_{m_r} (r = 1, 2, \cdots, D)$, where D is the number of camera orientations. The unknown motion is also projected as g_m within the global eigenspace. The most similar

motion is obtained from the global eigenspace using *Euclidian distance function* in (6).

$$d_m = \min_r \|\mathbf{g}_{m_r} - \mathbf{g}_m\| \tag{6}$$

6. Experimental Results.

6.1. Experimental setup. The experiments are performed with different synthesized human avatars performing ten different types of motions, namely *bend* (bending down), *carry* (carrying a box), *jump* (hopping in a place), *pjump* (jumping with two hands up and landing down), *pickup* (picking up something from the ground), *sitdown* (sitting down on a chair), *standup* (standing up from a chair), *stomachache* (touching stomach with pain and crouch), *walk* (walking motion), and *wave2* (waving two hands up in the air). The capture is varied by subject's height and shape, speed of motion, and field of view. The scene is assumed to be a backgroundless scene. Eight uncalibrated cameras are placed surrounding the avatar at 45 degrees apart, having 0-, 45-, 90-, 135-, 180-, 225-, 270- and 315-degree camera orientations. Figure 3 illustrates different motions, and the corresponding MHIs and XOR images. The *motion dataset* consists of 800 motion data separated into eight orientations. Among those, 560 motion data are considered as training set, and the rest as testing set. Thus the directional sub-databases consist of 70 motion data each. Likewise, the test set consists of 30 motions each (10 motions performed by 3 actors) for every orientation.

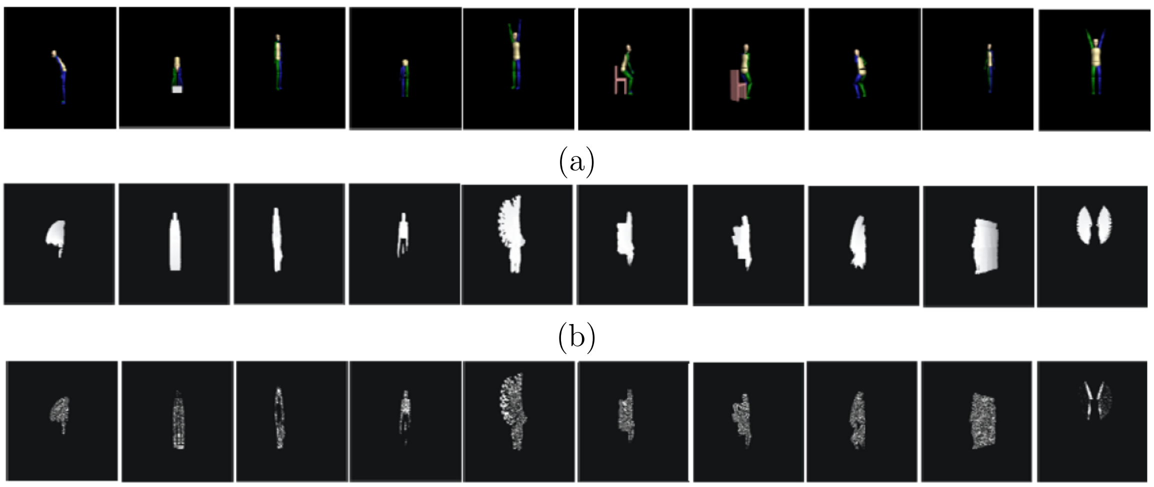


FIGURE 3. Motion representations from different viewpoints: (a) A single motion frame, (b) Corresponding MHIs, (c) Corresponding XOR images

6.2. Bin length. According to Section 4.1, the bin is defined as the hypercube within an eigenspace by portioning each eigen-axis. By transforming the extent of each axis to units with S divisions along the axis, the length of each edge of a bin is defined as:

$$\text{Bin Length}(L) = \frac{\text{Extent of an eigen-axis}}{\text{Total number of divisions along the axis}} = \frac{1}{S}$$

In the experiments, we shall compute the results by varying S and present the results in both tabular and graphical forms.

6.3. **Analysis of experimental results.** The modified structured motion database with the directional organization is used to evaluate the performance of the proposed approach. The experimental results are obtained by varying the division parameter that we call bin length. The results are extensively compared with the methods proposed in [16,18] which we refer to as *existing methods/strategies*.

Using MHI, in the case of no division, the recognition rates for proposed and existing strategies are 97% and 96%, respectively. The average recognition rates for the bin lengths 1/2 to 1/10 employing proposed and existing strategies are 86% and 74%, respectively; while the maximum recognition rate is achieved in our proposed approach for bin length 1/5 as 89%. So, we notice significant difference in recognition rates between the proposed and other database organizations. Similarly, using XOR images, in the case of no division, the recognition rate is 95% for the both cases. The average recognition rates for proposed and existing strategies are 77% and 64%, while the maximum recognition rate is achieved in our proposed approach for bin length 1/8 as 85%. Therefore, our proposed approach claims significant improvement over others for both the representations. The above results are tabulated in Table 1. From Table 1, it is noticeable that MHI outperforms XOR representation by considerable amount in terms of recognition rate.

Table 1. Recognition rate (%) for proposed and existing organization

Bin Length	**MHI**		**XOR**	
	Existing	*Proposed*	*Existing*	*Proposed*
1	96	97	95	95
1/2	80	88	58	75
1/3	76	88	61	73
1/4	80	82	67	78
1/5	68	89	63	76
1/6	78	83	68	75
1/7	70	85	64	73
1/8	75	83	70	85
1/9	70	86	69	79
1/10	65	86	60	78
Average	74	86	64	77

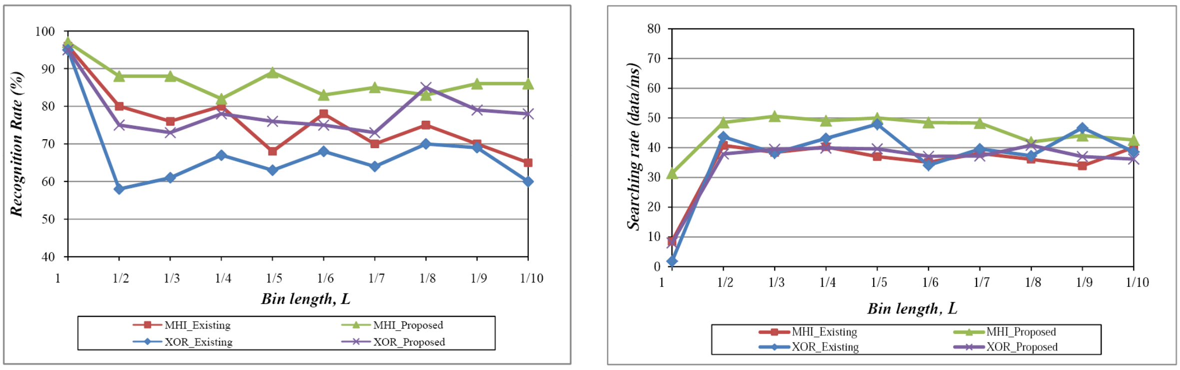
Moreover, we have also investigated the time requirement for the proposed method and existing methods. In the case of MHI with no division, the searching rates are 31.4 data/ms and 8.5 data/ms for proposed and existing organization, respectively. The average searching rates using the structured motion database concept are 51.4 data/ms and 37.8 data/ms, respectively. Therefore, using MHI the proposed approach is very much faster than the earlier approach. Similarly, using XOR images, in the case of no division, the searching rates are 8 data/ms and 1.8 data/ms for proposed and existing organization, respectively. The average searching rates using structured database concept are 39 data/ms and 41 data/ms, respectively. In the case of XOR, time requirement does not vary due to the high dimensionality of the feature space, while the proposed approach is three to four times faster than the sequential search. Table 2 illustrates the searching time and searching rates for both the aforesaid cases.

We notice that for every motion the average recognition rate for the proposed approach is considerably higher than the existing approach. Moreover, the comparative performances w.r.t. recognition rate and searching rate are also shown graphically in Figure 4. Figure 4(a) illustrates the supremacy of our proposed approach in terms of recognition

Table 2. Time consideration for proposed and existing organization

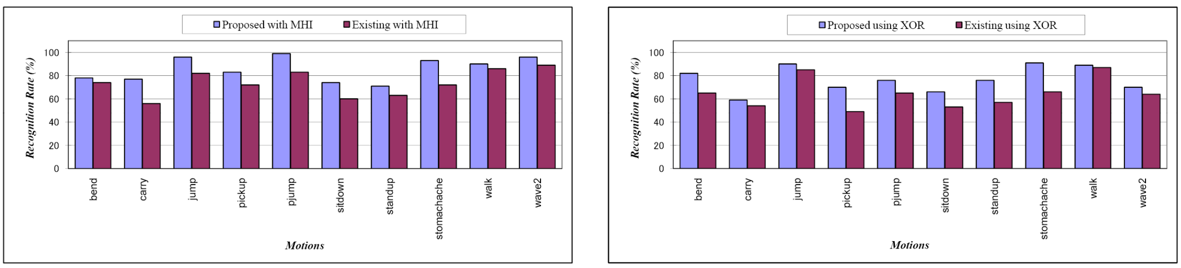
Bin Length	MHI				XOR			
	Existing		*Proposed*		*Existing*		*Proposed*	
	Searching Time(ms)	*Searching Rate*(data/ms)	*Searching Time*(ms)	*Searching Rate*(data/ms)	*Searching Time*(ms)	*Searching Rate*(data/ms)	*Searching Time*(ms)	*Searching Rate*(data/ms)
1	66.17	8.5	17.83	31.4	308.55	1.8	70.16	8
1/2	13.72	40.8	8.9	62.9	12.81	43.7	14.625	38.3
1/3	14.6	38.4	9.54	58.7	14.66	38.2	13.825	40.5
1/4	13.89	40.3	9.55	58.6	13	43.1	13.9	40.3
1/5	15.15	37	10.07	55.6	11.69	47.9	13.95	40.1
1/6	15.96	35.1	11.55	48.5	16.4	34.1	15.28	36.6
1/7	14.68	38.1	11.57	48.4	14.15	39.6	15.18	36.9
1/8	15.51	36.1	13.36	41.9	15.02	37.3	13.72	40.8
1/9	16.5	33.9	12.64	44.3	12	46.7	14.175	39.5
1/10	13.97	40.1	12.73	44	14.5	38.6	14.82	37.8
Average	14.9	37.8	11.1	51.4	13.8	41	14.4	39

rates (%) for MHI and XOR images. From these figures, we can figure out the most acceptable case in terms of recognition rate and searching time. According to this, it is found that the performance of the system adopted in [16,18] can be largely improved by using MHI with the proposed organization. Figure 5 illustrates the motion-wise performance improvement for the proposed approach. The experiments were conducted on a 2.93 GHz Processor 4GB RAM-computer.



(a) (b)

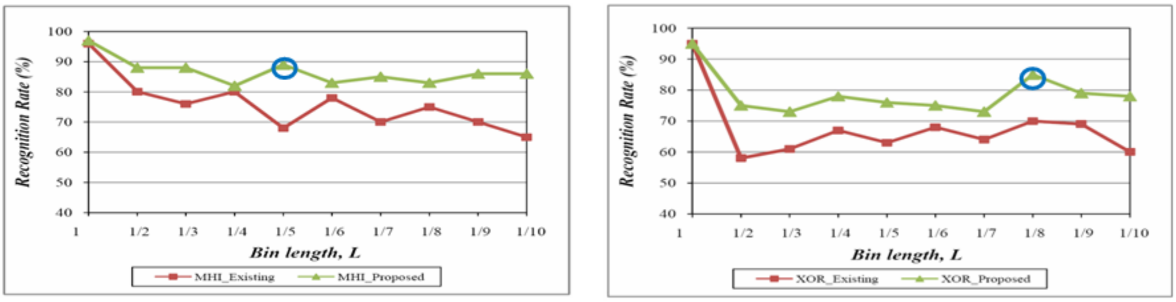
Figure 4. Comparative performance for *proposed* versus *existing* methods using MHI and XOR images; (a) Recognition rates, (b) Searching rates



(a) (b)

Figure 5. Motion-wise performance improvement; (a) MHI, (b) XOR images

7. **Discussion.** The experimental results signify the importance of the directional organization of the motion data in the case of various types of motions with several orientations. Due to the registration of directionally similar, but originally different, motions, confusion arises. Our proposed directional approach is an adaption of the problem of similarity searching. It is experimentally shown that the recognition rate for the conventional approach proposed in [16,18] performs poorly when the types of motions and the number of orientations increase. Although, we notice that XOR-based recognition takes slight longer searching time than that of MHI, it is surely acceptable for the recognition compared to prior researches. Therefore, we can claim that both MHI and XOR, without taking into account of local motion dynamics, show satisfactory performance for our proposed system. However, from Figure 6, we get some idea about the proper bin length for the system using either MHI or XOR image. The bin length is to be selected in such a way to optimize the system's performance in terms of recognition rate and searching time. For our experimental dataset, we can choose the length of 1/5 for MHI- and 1/8 (as shown in *blue circle* in Figure 6) for XOR image-based system according to recognition rate by compromising with time requirement.



(a) (b)

FIGURE 6. Possible selection of *bin length*; (a) MHI, (b) XOR images

In the performed experiments, the motions were formed by varying speed, and varying shape of the performing subjects. In spite of these variabilities, our proposed motion database with directional organization shows satisfactory performance on their recognition. In Section 4.3, the robustness of this approach is analyzed and the reasons behind the superiority of the method over non-directional organization are also represented by logical inference. The occurrence of the motions, i.e., *how* the motion is moving, is another cause of poor recognition. The experimental motions, however, are not so much complex, rather simple and have almost no overwriting problems. For the overwriting cases, the Directional MHI representation [23] has much potential to uniquely represent each motion. This form of motion template may also be incorporated in our proposed system.

8. **Conclusions and Future Work.** We proposed a novel approach for human motion recognition by developing an efficient motion database that is organized based on camera orientation. The directional data organization strategy is adopted here to organize and manage the motion database. Experiments were performed in which the motions are performed by synthesized actors; those exclude the noise factor where the noise or background may highly affect the system's performance. The effectiveness of the proposed organization was proved by analyzing the experimental results. We found that directional organization with MHI gained about 90% recognition rate in comparison with other recognition approaches. This proves the excellence of the proposed approach.

Although we have achieved satisfactory performance with our proposed strategy, the accuracy of the system is still about 90% which is subjected to be improved to the results of the blind search approach. This deficiency of the results is likely to be overcome by modifying the *Nearest Neighbor Searching* algorithm to accurately search the most similar motions corresponding to the unknown motions. The approach recently proposed in [24] might be employed for the problem resolution. Moreover, the real-life indoor and outdoor motions should be used for the practical application of the system. Separate mechanism for background subtraction and foreground segmentation might be employed for real-life motion recognition cases. Moreover, the investigation is vitally required to automatically select the division parameter that corresponds to the best performance. Motion dataset-based analysis may lead to the appropriate selection of the parameter.

Above all, except some possible modifications, our system works well with various kinds of motions with different camera orientations. The proposed approach is certainly a one-step further in developing intelligent robots to instantly recognize different human motions in real world environment. The applicability of the system might be increased by further enhancements.

Acknowledgment. This work was supported by (JSPS) KAKENHI (22510177) Grant-in-aid for Scientific Research (C).

REFERENCES

[1] J. K. Aggarwal and Q. Cai, Human motion analysis: A review, *Computer Vision and Image Understanding*, vol.73, no.3, pp.428-440, 1999.
[2] J. Blackburn and E. Ribeiro, Human motion recognition using isomap and dynamic time warping, *Human Motion Understanding, Modeling, Capture and Animation, LNCS*, pp.285-298, 2007.
[3] D. M. Gavrila, The visual analysis of human movement: A survey, *Computer Vision and Image Understanding*, vol.73, no.1, pp.82-98, 1999.
[4] O. Masoud and N. Papanikolopoulos, A method for human action recognition, *Image and Vision Computing*, vol.21, no.8, pp.729-743, 2003.
[5] L. Wang and D. Suter, Analyzing human movements from silhouettes using manifold learning, *Proc. of IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pp.7, 2006.
[6] A. Elgammal and C. S. Lee, Inferring 3D body pose from silhouettes using activity manifold learning, *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp.681-688, 2004.
[7] H. Fujioka, H. Kano and X. Chen, Motion recovery under perspective stereo vision, *International Journal of Innovative Computing, Information and Control*, vol.5, no.1, pp.167-182, 2009.
[8] D. Gavrila and L. Davis, 3-d model-based tracking of human upper body movement: A multi-view approach, *Proc. of Int. Symposium on Computer Vision*, pp.253-258, 1995.
[9] M. Isard and A. Blake, Condensation – Conditional density propagation for visual tracking, *Int. J. Computer Vision*, vol.29, no.1, pp.5-28, 1998.
[10] T. B. Moeslund, A. Hilton and V. Kruger, A survey of advances in vision-based human motion capture and analysis, *Computer Vision and Image Understanding*, vol.104, no.2, pp.90-126, 2006.
[11] T. Ogata, J. K. Tan and S. Ishikawa, High-speed human motion recognition based on motion history image and an eigenspace, *IEICE Trans. Information and Systems*, vol.E89-D, no.1, pp.281-289, 2006.
[12] T. B. Moeslund and E.Granum, A survey of computer vision-based human motion capture, *Computer Vision and Image Understanding*, vol.81, pp.231-268, 2001.
[13] M. A. R. Ahad, T. Ogata, J. K. Tan, H. S. Kim and S. Ishikawa, A complex motion recognition technique employing directional motion templates, *International Journal of Innovative Computing, Information and Control*, vol.4, no.8, pp.1943-1954, 2008.
[14] A. F. Bobick and J. W. Davis, The recognition of human movement using temporal templates, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.23, no.3, pp.257-267, 2001.
[15] X. He and P. Niyogi, Locality preserving projections, in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. K. Saul and B. Schölkopf (eds.), MA, MIT Press, 2004.
[16] J. K. Tan, K. Kouno, S. Ishikawa, H. Kim and T. Shinomiya, High-speed human motion recognition employing a motion database, *J. Institute of Image Electronics Engineers of Japan*, vol.36, no.5, pp.738-746, 2007.

[17] R. Bayer and E. McCreight, Organization and maintenance of large ordered indexes, *J. Acta Informatica*, vol.1, no.3, pp.173-189, 1972.

[18] J. K. Tan and S. Ishikawa, High accuracy and real time recognition of human activities, *Proc. of IEEE Industrial Electronics Society Conf.*, pp.2377-2382, 2007.

[19] H. Murase and S. K. Nayar, Visual learning and recognition of 3-D objects from appearance, *Int. J. Computer Vision*, vol.14, no.1, pp.5-24, 1995.

[20] S. M. A. Eftakhar, J. K. Tan, H. Kim and S. Ishikawa, Human motion recognition employing large motion-database structure, *Int. J. Advanced Computer Engineering*, vol.2, no.1, pp.17-23, 2009.

[21] D. J. Bernd and J. Clifford, Finding patterns in time series: A dynamic programming approach, *Advances in Knowledge Discovery and Data Mining*, pp.229-248, 1996.

[22] R. D. Green and L. Guan, Quantifying and recognizing human movement patterns from monocular video images-part I: A new framework for modeling human motion, *IEEE Trans. Circuits System Video Technology*, vol.14, no.2, pp.179-190, 2004.

[23] M. A. R. Ahad, T. Ogata, J. K. Tan, H. S. Kim and S. Ishikawa, Analysis of motion self-occlusion problem due to motion overwriting for human activity recognition, *J. Multimedia*, vol.5, no.1, pp.36-46, 2010.

[24] S. M. A. Eftakhar, J. K. Tan, H. Kim and S. Ishikawa, Improvement of a structured motion database for high accuracy human motion recognition, *Int. J. Biomedical Soft Computing and Human Sciences*, vol.17, no.1, pp.19-28, 2011.