# A Study on N-Spherical Histogram for Unbounded Tensor Features in Human Action Recognition

Ngoc Nam Bui, Jin Young Kim
Dept. of Electronics and Computer Engineering Chonnam National University 500-757
Yongbong-dong, Buk-gu, Gwangju, South Korea
buingocnam87@gmail.com, beyondi@chonnam.ac.kr

**Abstract: - Recent years, Dense Trajectory features has a crucial role in extracting implicit features for action recognition. The method encloses motion and appearance descriptors to specify characteristics of each trajectory. Moreover, combining gradient and optical flow field using tensor product has made a strong positive impact on the result as we introduced in our previous work. In this paper, a breakthrough concept of encoding a high dimensional unbound space using spherical coordinate is introduced and imposed to obtain sophisticated spherical tensor features. The experimental result shows that our propose features outperforms other conventional ones and the combination of all feature channels achieves the highest accuracy rate in our selfrecorded dataset.**

**Key-Words: - n-spherical coordinate, tensor features, GMM-UBM supervector, Human Action Recognition, multidimensional histogram.**

## I. INTRODUCTION

Recognizing human actions is always challenged due to the source of information contains much details as well as redundancies. Since the early dates, people attempts to estimate motion characteristic through the displacement of pixels in a sequence of images. Silhouette image is widely used at the beginning due to its efficiency in revealing target boundary and the motion flow. Specifically,Davis and Bobbick [1] in their work has transformed the Silhouette image into more sophisticated features named Motion Energy Image (MEI) and Motion History Image (MHI) which respectively express the shape and temporal motion history. According to [2], the global information obtained from Silhouette Image is highly sensitive to noise. Wang, et al. [3] proposed a novel method by extracting a flexible cubic following an optical trajectory. Then, different types of descriptor are extracted and passed through a Bag Of Feature Words for encoding purpose. In [4], we argue that the GMM-UBM supervector can obtain a better result instead of BOF or Fisher Vector [5]. Later on, a tensor features [6] which targets to modify the combination of gradient and optical field at low level has been introduced and shown their encouraged result in term of KTH open dataset. In the scope of this paper, we instead of projecting a 4d tensor into 2 dimensional space, we introduce a multidimensional histogram based spherical coordinate.

We organize our paper as follows. In Section 2, the dense trajectory is revised briefly with the introduction of tensor combination. Then, our main focus, the spherical histogram, is analyzed and explained fruitfully in section 2.1. The overall structure of HAR systm based GMM-UBM supervector is mentioned in Section 3. We verify our propose method using self-constructed dataset CNU and shows the results in Section 4. Finally, Section 5 is in charge of concluding our work with all the advantages as well as shortcomings.

## II. MULTIDIMENSIONAL UNBOUNDED SPACE HISTOGRAM OR SPHERICAL HISTOGRAM.

### *A. Dense Trajectory and Tensor features.*

Dense Trajectory features [3] by extracting descriptors across along a trajectory has demonstrated its best performance over various datasets. Specifically, 5 types of descriptor including Position, Histogram of Gradient (HOG), Histogram of Flow (HOF) and Motion Boundary Histogram (MBH) are utilized to represent trajectory's characteristics. In our previous work [6], we developed a tensor combination based optical flow and image gradient quantity that seems efficient to reveal essential motion properties.

Particularly, vectors obtained by gradient and optical flow should first be integrated into a single entity that is either a scalar or another vector, and then encoded into a descriptor. Furthermore, the proposed feature is oriented to express what kind of motion is influenced by the appearance model and vice versa. In other words, the X flow and Y flow motion is affected by the gradients X and Y, which is similar to the case of stress tensors in continuum mechanics. Given a finite set, a vector space of $\{\boldsymbol{g},\boldsymbol{f}\}$ (which stands for gradient and flow vector in our case), a tensor is defined as follows:

$$T\{\boldsymbol{g},\boldsymbol{f}\} = \boldsymbol{g} \otimes \boldsymbol{f} = \boldsymbol{f} \otimes \boldsymbol{g}, \quad (1)$$

where $T$ is denoted for tensor and $\otimes$ is the tensor product or dyadic product. Then the tensor matrix and divergence tensor is applied to reduce the dimensions of features to serve for encoding purpose.

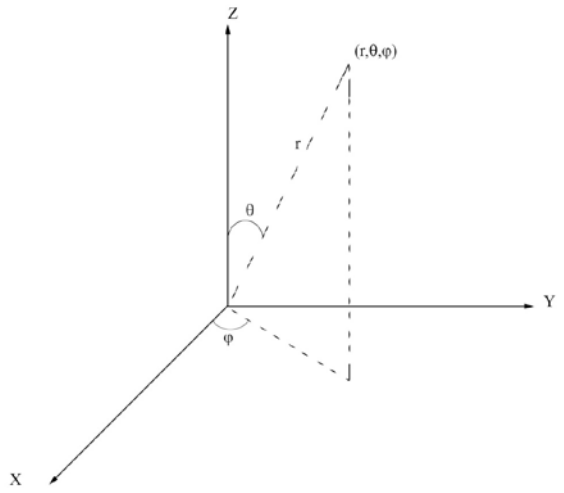


Figure 1: The correlation between $\mathbb{R}^3$ and 3-sphere coordinate.

*B. Multidimensional Unbounded Space Histogram.*

Although, the tensor feature has outperformed the others as in [6], forcing a 4d tensor into 2d vector compresses much information which somehow degrades the performance. Therefore, we attempt to keep the original dimension of our tensor features and develop a method to estimate their distribution.

Histogram is a bona fide probability estimation that developed early by Pearson [7]. In order to obtain the estimation, the histogram first "bins" the range of data into smaller intervals. Then each of frequency's interval is measured by the value of those has fallen into. It implies the bounded condition for sample values; otherwise, number of bins would be unlimited. Unfortunately, unlike intensity values which has the range 0 to 255, tensor features belongs to 4 dimensional unbounded domain. Therefore, a preprocessing mechanism is required to set a range limit for these features. Motivating by the Histogram of Orientation, we apply the projection from $\mathbb{R}^n$ to n-sphere coordinate as illustrated in Figure 1 for the case $n = 3$. Given a vector $\boldsymbol{x} \in \mathbb{R}^n$, the mathematical formula for the mapping onto spherical space is given as follows

$$\kappa_n : \mathbb{R}^n \to \mathbb{R}x[0,360]^{n-1}$$
$$\boldsymbol{x} = (x_1, x_2, \dots, x_n) \to s = (r, \theta_1, \dots, \theta_n), \quad (2)$$

where $r$ and $\theta_i$ represents for the radial and angular coordinates as derived by Equation (3).

$$r = \sqrt{x_n^2 + x_{n-1}^2 + \cdots + x_1^2}$$
$$\theta_1 = arccot \frac{x_1}{\sqrt{x_n^2 + x_{n-1}^2 + \cdots + x_2^2}}$$
$$\theta_2 = arccot \frac{x_2}{\sqrt{x_n^2 + x_{n-1}^2 + \cdots + x_3^2}} \quad (3)$$
$$\theta_{n-2} = arccot \frac{x_{n-2}}{\sqrt{x_n^2 + x_{n-1}^2}}$$
$$\theta_{n-1} = 2arccot \frac{x_{n-1} + \sqrt{x_n^2 + x_{n-1}^2}}{x_n},$$

where if $x_k \neq 0$ but $x_{k+1} = \cdots = x_n = 0$ then we check the following condtion to assign value for $\theta_k$

$$h_{b_{i_1} b_{i_2} \dots b_{i_{n-1}}}\left(\theta_{x,y}^1 \theta_{x,y}^2 \dots \theta_{x,y}^{n-1}\right) = h_{b_{i_1} b_{i_2} \dots b_{i_{n-1}}}\left(\theta_{x,y}^1 \theta_{x,y}^2 \dots \theta_{x,y}^{n-1}\right) + \mathrm{r} \quad (4)$$
$$if \quad \left\lfloor \theta_{x,y}^1 \right\rfloor \in b_{i_1}, \quad \left\lfloor \theta_{x,y}^2 \right\rfloor \in b_{i_2}, \dots, \left\lfloor \theta_{x,y}^{n-1} \right\rfloor \in b_{i_{n-1}}$$

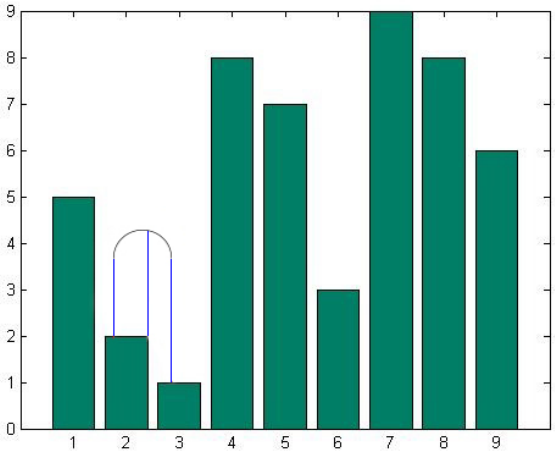


Figure 2: *The existing problem of hard binning scheme.*

The hard assignment above has leaves out the residuals to be undefined that is shown in Figure 2. The actual angle can be observed as a half circle which spreads out among two bins. The, according to Equation (4), the whole values of r will be summed up to only the left bin whereas it should be shared with the right bin also.

The residual grows exponentially with dimensions as in Figure 3 that causes the estimated distribution to be far different from the original one. In three dimensional spaces, the arc is represented by a half sphere that portioned its value toward four consecutive bins. The red volume indicates the bin which occupies the most portion of a given angular vector. Obviously, the hard binning assignment will add only the dominant one with the whole amount of r. We therefore propose a smoother assignment by considering the portion of the angular as formalized in the equation below.

$$r_{re}^{j} = \left(\theta_{x,y}^{j} - \left\lfloor \theta_{x,y}^{j} \right\rfloor\right) * \mathrm{r}$$

$$if \qquad \left\lfloor \theta_{x,y}^{j} \right\rfloor \in b_{i_j}, j = \overline{1, n-1}$$

$$h_{b_{i_{1+q}} \ldots b_{i_j+1} \ldots b_{i_{n-1+q}}}\left(\theta_{x,y}^{j}\right) = \tag{5}$$

$$h_{b_{i_{1+q}} \ldots \left[\left(b_{i_j}+1\right)\%nBin\right] \ldots b_{i_{n-1+q}}}\left(\theta_{x,y}^{j}\right) + r_{re}^{j}/\mathrm{V}$$

$$h_{b_{i_{1+q}} \ldots b_{i_j} \ldots b_{i_{n-1+q}}}\left(\theta_{x,y}^{j}\right) = h_{b_{i_{1+q}} \ldots b_{i_j} \ldots b_{i_{n-1+q}}}\left(\theta_{x,y}^{j}\right) + \left(\mathrm{r} - r_{re}^{j}\right)/\mathrm{V},$$

with $\mathrm{V} = \frac{2^{\mathrm{n}-1} * (\mathrm{n}-1)}{2}$ and $q = [0,1]$.

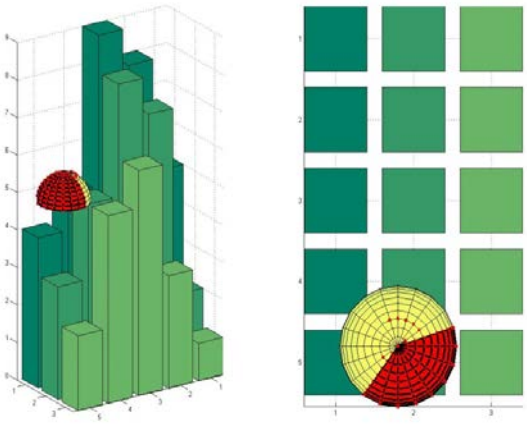

Figure 3: *The existing problem of hard binning scheme in 3 dimensional spaces.*

## III. SYSTEM STRUCTURE.

The block diagram of our system based on Dense Trajectory Features with GMM-UBM supervector is illustrated in Figure 4.

a) At first, the dense trajectories features with various descriptors including our proposed spherical tensor representation are extracted and preprocessed using Principal Component Analysis (PCA).

In the next stage, we utilize the GMM-UBM method. All features are pooled to train the GMM-UBM model by utilizing an Expectation Maximization (EM) algorithm.

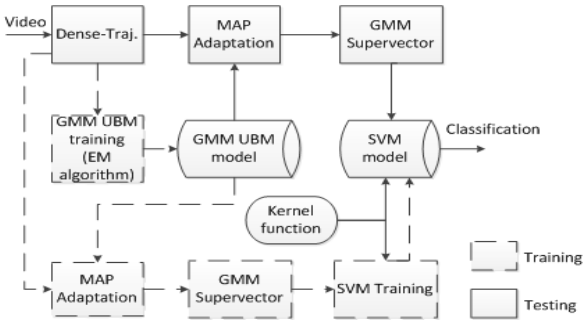


Figure 4: The block diagram of the human action recognition system.

b) After that, a GMM for each sample is derived by using a Maximum a Posterior (MAP) adaption with the UBM model.

c) A mean supervector is obtained from the GMM model by cascading all Gaussian components into a single vector.

d) Finally, the SVM classifier is trained with different kernel functions. We adopted the nonlinear GMM KL and GUMI as a primary kernel due to their efficiency that has been verified in our previous work [4].

## IV. DATASET AND EXPERIMENTAL RESULTS.

### *A. Dataset.*

We recorded our data set focusing on capturing violent behavior in school. The data collection has been conducted at our university and surrounding areas while ensuring the realistic of human under various types of scenarios. For each condition, we collected the information for two groups, males and females, performing various actions. Moreover, the school environment requires our method to overcome the similarity causing by student uniform. In several groups such as drinking fountain, the scale of targets is diversified dramatically. We divide our training and testing set as follows:

- 1091 fighting scenes with 665 ones are for training and the remaining is testing data.

• 1268 non-fighting scenes with 809 videos are used to train the data and the rest are provided to evaluation purpose.


Figure 5*: Sample clips captured from CNU dataset.*

*B. Experimental Results*.

We experiment the CNU dataset with multidimensional sphere histogram respects to tensor features and the combination [MBHx MBHy], which is also a 4 dimensional vector. In Table 1, we show our experiment of SphereTensor feature which obtained the best accuracy using 4 mixture GMM model. Moreover, we integrate the new features with the five conventional descriptors and depict our result in Table 2. In the case of GUMI kernel, the combination with Tensor Matrix, Divergence Tensor and Vorticity Tensor features give highest score with 77.49%. Otherwise, using GMM KL kernel, the combination of SphereTensor and others increase the accuracy rate to 1.63% that also is the best one so far.

*Table 1: Comparison SphereTensor features with others.*

| **4 mixtures** | | | |
|---|---|---|---|
| | **Linear** | **Nonlinear GUMI** | **Nonlinear GMMKL** |
| Traj | 53.25 | 66.01 | 64.5 |
| HOG | 61.83 | 71 | 70.65 |
| HOF | 64.97 | 71.46 | 72.16 |
| MBHx | 62.06 | 71.23 | 71.58 |
| MBHy | 66.36 | 71.93 | 71.23 |
| Tensor | 58.12 | 68.21 | 68.1 |
| Divtensor | 58.47 | 67.63 | 68.21 |
| VortTensor | 62.88 | 67.87 | 65.43 |
| SphereTensor | 62.65 | 71.23 | 72.27 |
| **8 mixtures** | | | |
| | **Linear** | **Nonlinear GUMI** | **Nonlinear GMMKL** |
| Traj | 57.42 | 66.94 | 67.63 |
| HOG | 64.62 | 70.42 | 70.42 |
| HOF | 68.45 | 73.43 | 74.13 |
| MBHx | 65.31 | 73.09 | 73.78 |
| MBHy | 65.2 | 72.27 | 72.16 |
| Tensor | 61.72 | 68.21 | 66.36 |
| Divtensor | 64.15 | 69.84 | 69.95 |
| VortTensor | 61.6 | 67.52 | 67.75 |
| SphereTensor | 64.97 | 73.43 | 73.43 |

The multidimensional spherical histogram can be applied to any unbounded vectors. Therefore, we cascade the two MBH features to form a single one which has 4 components and apply the technique to estimate their distribution. In the sense of individuals feature, the SphereMBH shows much better result that compare to other conventional ones

as shown in Table 3. However, when combining them together in Table 4, the performance is downgraded.

*Table 2: Comparison SphereTensor features with others.*

| **4 mixtures** | | | |
|---|---|---|---|
| | **Linear** | **Nonlinear GUMI** | **Nonlinear GMMKL** |
| (Traj.+HOG+HOF+MBHx+MBHy) | 71.69 | 76.68 | 76.8 |
| (Traj.+HOG+HOF+MBHx+MBHy)+(Tensor+DivTensor+VortTensor) | 71.35 | 76.45 | **77.73** |
| (Traj.+HOG+HOF+MBHx+MBHy)+SphereTensor | **72.39** | **76.91** | 77.49 |
| **8 mixtures** | | | |
| | **Linear** | **Nonlinear GUMI** | **Nonlinear GMMKL** |
| (Traj.+HOG+HOF+MBHx+MBHy) | 71.69 | 77.49 | 77.61 |
| (Traj.+HOG+HOF+MBHx+MBHy)+(Tensor+DivTensor+VortTensor) | 72.62 | 77.49 | 78.19 |
| (Traj.+HOG+HOF+MBHx+MBHy)+SphereTensor | 70.07 | 76.91 | **79.12 (best)** |

*Table 3: Comparison SphereMBH features with others.*

| **4 mixtures** | | | |
|---|---|---|---|
| | **Linear** | **Nonlinear GUMI** | **Nonlinear GMMKL** |
| Traj | 53.25 | 66.01 | 64.5 |
| HOG | 61.83 | 71 | 70.65 |
| HOF | 64.97 | 71.46 | 72.16 |
| MBHx | 62.06 | 71.23 | 71.58 |
| MBHy | 66.36 | 71.93 | 71.23 |
| SphereMBH | **69.95** | **75.52** | **76.57** |
| **8 mixtures** | | | |
| | **Linear** | **Nonlinear GUMI** | **Nonlinear GMMKL** |
| Traj | 57.42 | 66.94 | 67.63 |
| HOG | 64.62 | 70.42 | 70.42 |
| HOF | 68.45 | 73.43 | 74.13 |
| MBHx | 65.31 | 73.09 | 73.78 |
| MBHy | 65.2 | 72.27 | 72.16 |
| SphereMBH | **69.03** | **73.55** | **73.9** |

*Table 4: Comparison SphereTensor features with others.*

| **4 mixtures** | | | |
|---|---|---|---|
| | **Linear** | **Nonlinear GUMI** | **Nonlinear GMMKL** |
| (Traj.+HOG+HOF+MBHx+MBHy) | 71.69 | 76.68 | **76.8** |
| (Traj.+HOG+HOF+MBHx+MBHy)+SphereMBH | 71.46 | 76.57 | 76.68 |
| **8 mixtures** | | | |
| | **Linear** | **Nonlinear GUMI** | **Nonlinear GMMKL** |
| (Traj.+HOG+HOF+MBHx+MBHy) | 71.69 | 77.49 | **77.61** |
| (Traj.+HOG+HOF+MBHx+MBHy)+SphereTensor | 71.81 | 76.8 | 77.49 |

## V. CONCLUSION

We propose a critical solution to handle the encoding problem of multidimensional unbounded data and experiment on the tensor matrix. The spherical encoding has demonstrated its outstanding performance compare to HOG, HOF or MBH features in HAR system. Moreover, our concept can be adapted to different fields that require estimating a distribution of high dimensional space.

## ACKNOWLEDGEMENT

This research was supported by the Ministry of Education, Science Technology (MEST) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation.

## REFERENCES

[1] J. W. Davis and A. F. Bobbick, "The Representation and Recognition of Action Using Temporal Templates," IEEE Trancsactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 257-267, March 2001.

[2] R. Poppe, "A survey on vision-based human action recognition," Image and Vision Computing, vol. 28, no. 6, pp. 976-990, June 2010.

[3] H. Wang, A. Kläser, C. Schmid., and C.-L. Liu, "Dense Trajectories and Motion Boundary Descriptors for Action Recognition," International Journal of Computer Vision, vol. 103, no. 1, pp. 60-79, 2013.

[4] N. N. Bui and Y. J. Kim, "Human action recognition based on GMM-UBM supervector using SVM with non-linear GMM KL and GUMI," in SPIE 9631, Seventh International Conference on Digital Image Processing (ICDIP 2015), July 2015.

[5] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the Fisher Kernel for Large-Scale Image Classification," in Computer Vision – ECCV, pp. 143-156, January 2010.

[6] N. N. Bui, J. Y. Kim, and H.-G. KIM, "Gradient-Flow Tensor Divergence Feature for Human Action Recognition," Journal IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. E99-A, no. 1, pp. 437-440, January 2016.

[7] K. Pearson, "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 186, no., pp. 343-414, 1895.