Kernel-Based on Data Fusion for Image Classification with Body Energy Action Model

Nutchanun Chinpanthana
Faculty of Information Technology, Dhurakij Pundit University, 110/1-4 Prachachuen Rd., Laksi, Bangkok 10210, Thailand
Email: nutchanun.cha@dpu.ac.th

Tejtasin Phiasai
Faculty of Engineering, King Mongkut's University of Technology Thonburi, Pracha-utid Rd., Bangmod, Bangkok 10140, Thailand
Email: tejtasin@gmail.com

***Abstract*—Human action classification has been and still a highly interesting and important research topic. The research results are capable of analyzing human action that we visually perceive in many aspects. Therefore the research requires an effective and competent approach to accurately interpret human action. In this paper, we present a novel model called the body energy action model for finding actual semantic action. The model is based on the fundamental concepts of biomechanics that human movement in different classes is likely to spend different amounts of energy. The model is classified by using kernel-base data fusion to obtain from the 5-fold cross validation. Experimental results show that the proposed provides much more authentic meaning of human actions.**

***Index Terms*—human action, classification, semantic images, image classification; kernel function**

I. Introduction

Recently, semantic classification has been playing an important role in human image research, since the digital camera technology has been extremely successful developed over the last decade [1], [2], [3]. People can take photos quickly and easily, so the number of digital images has increased significantly. When people need to find a desired image, they often spend time searching the images in a large database. Researchers attempt to make various methods for semantic classification. Keyword technique is one of approach that is used in classification [4], [5], [6]. The results have successfully competed for matching in the term of keywords but do not make sense in the term of human meaning. For this reason, the classification results are not directly to the semantic human images. Human action has become an important field of research [7], [8]. We briefly discuss on some significant examples and categorize them into two main research directions: Automatic image annotation approaches and Human action based approaches.

Manuscript received August 1, 2013; revised November 22, 2013.

A. Automatic Image Annotation Approaches

There are ongoing researches [6], [9], [10] in an automatic image annotation system which has attempting to discriminate the recognition low-level feature extraction process by labeling the objects with keywords. Important elements in an image are manually labeled. These labeled are called keyword. Then, every image in the database is compared against those keywords to detect the specific keywords of the image. The methods are developed and used for extracting semantic consists of words form dictionary. The concept of word relationships is used for clustering images. Some researchers have addressed the issues of learning of term similarity matrix and word grouping for intelligent query expansion. They construct more meaningful concept clusters of co-occurring keywords technique. For example, a user needs to find an image "a man resting on the beach". The irrelevant images that are labeled with a set of beach keywords are also returned. Results still remain unsatisfactory. To extend the online annotation tools [11], [12], LabelMe [13], [14], PoseShop [2] and Ellen Molitorisová [5] presented a data set with region-based annotations. There are various methods to execute region-based annotations such as bounding box, polygonal, and freehand drawing. The disadvantage is that it takes time to draw a line around object. In LabelMe [13], [14], the annotation are linked to WordNet [15], [16], [17], a lexical database in which nouns, verbs, adjectives, and adverbs are organized into a sets based on their keyword meanings. An effort is made to find the correct meaning in WordNet automatically. However, many words have more than one sense, which makes it hard to find the correct keyword with the correct meaning.

Although keywords have more direct mapping toward high-level semantics than low-level visual features, it does not represent whole image meaning. The set of keywords in the image is not related to the whole-semantics directly. Users have a desire image in mind as a sense of semantics but it does not a keyword sense. By reviewing each paper, the process does not concern on

doi: 10.12720/ijsps.1.2.218-224

activities from human image. The main character in personal photos is human that shows various postures. To overcome this, researchers combine the human actions that are applied to personal photographs into framework.

B. Human Action Based Approaches

Human action is purposeful behavior that integrates whole body movement to signify the hidden meaning of the mind [8], [11]. Developing algorithms to classify human actions has proven to be a challenge [12], [13]. Therefore, the recognition of human actions has become a task of high interest within the field [18], [19], especially for health-care, medical sector, military and automated surveillance. For instance, patients with diabetes, cardiovascular disease or heart disease are usually required to body comfort. They can adapt the results of the body action as a way of diagnosing patient's symptom. The recognizing their activities such as walking, running, lying, or sitting becomes useful task. In earlier work, researchers [20], [21] focus on human standard posture recognition: standing, sitting, and lying. Recognition of the basic action seems useful but it is limited to represent the details of action. Sukthankar and Sycara [22] recognized human movement by employing laws and concepts of representation through the use of the structure of directed acyclic graphs. This representation structure was classified by using support vector machines. The results produced action candidates (e.g., sneak, probe, crunch). Although the structure reveals the human movement based on graph more clearly, the semantic images were not investigating the intuitive meaning of such scenes. Other group researcher [5], [8] can be recognized fairly complex activities such as eating, taking a shower, washing dishes, etc. But, it rely on data from a number of sensors placed in target of objects which people are supposed to interact with such as golf, dishes, table, pencil, etc.

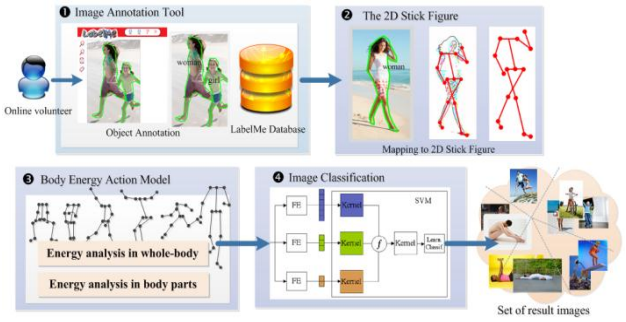

Figure 1. Overview the action classification process of body energy action model proposed approach

Most researchers interpret the image meaning from an outstanding human object which the primary thing is giving more information in the whole scene. People will observe purely superficial things about human object and make judgments based on those observations. Certainly the interpretation starts with what we wear, but it is also how we act. Therefore, we attempt to find the intuitive meaning of dominant human by analyzing the human action. Consequently, we integrate these concepts in order to analyze the semantics of human activities with the Body Action Energy Model, which adapts two research areas. Physical anthropology is the study the energy intensity of estimate metabolic expenditure in the human body, and Biomechanics is the study of mechanics and physical expenditure.

The objective of this work is to purpose the model of body energy action based on principle of Physical anthropology and the Biomechanics [23], [24]. The energy expenditure action is primary features which are extracted and calculated from proposed method. Our processing system is divided into 4 main steps: Image Annotation Tool, The 2D stick figure model, Body Energy Action Model and Image Classification as shown in Fig. 1.

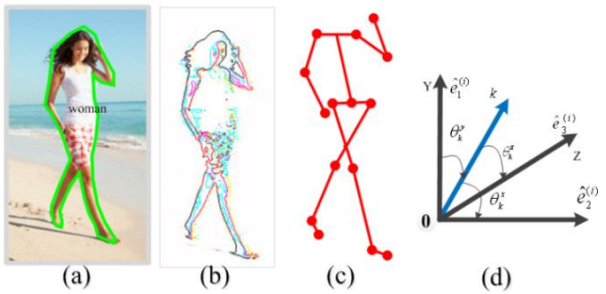

Figure 2. An annotated image (b) Detected contour body (c) Constructed to the stick figure model (d) Limb orientation is represented with the direction cosines of angles $(\hat{e}_1^{(i)}, \hat{e}_2^{(i)}, \hat{e}_3^{(i)})$ between the limb K and the axes.

The rest of paper is original as follows: annotating image and extracting information into the 2D stick figure model is presented in Section II. The computation of the body action energy model is detailed in Section III. Section IV describes the support vector machine and kernel function Classification. Experimental and evaluation results are presented in Section V. Section VI is the summary of the results and the possibility of future works.

II. Data Preprocessing

A. Image Annotation Tool

The first thing to consider within the proposed approach of human activity recognition is what information to use in an image database. We are concentrating on the meaning of the action and how we find the essential features. Based on our observation, a semantic image is emerging from major contents and the association between image contents. Each content has different types however some types possibly have similar semantics. For example, "stone" "boulder", "clay" and "rock" are semantically analogous because they are both instances of "rocky mountain". This relation might be proven useful in the part of semantically related contents. Hence, we selected LabelMe [13], [14] annotation tools that is evaluated with ontology and linked to WordNet [17]. The LabelMe database is the most comprehensive public image database that is manually annotated by

online volunteers as shown in Fig. 1. Therefore, we use appropriate images automatically annotated from LabelMe Tool that contains large scale image collections.

B. The 2D Stick Figure Model

In this paper, we focus on the human action. The body of human is a primary content for extracting information. Hence, we selected the dominant human images from LabelMe collections into the data set as input. We have investigated the use of the stick-figure via the bogy region extraction proposed by Jong-Seung Park [25]. It has many advantages to implement the stick-figure model with body boundary contours. Body boundary contours provide essential information to reconstruct a body model as shown in Fig. 2b. Each body part is considered as a 3D cylinder and its projection to an image plane is a 2D ribbon. Each ribbon is corresponding to body parts.

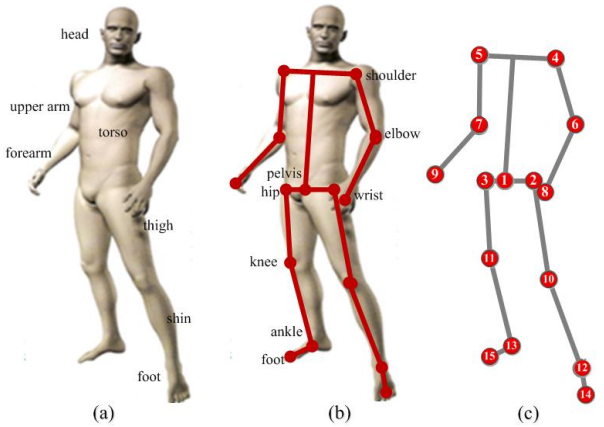

Figure 3. (a) The human model in standing posture (b) Skeleton mapping to body parts. (c) The reference point in the stick figure model

Fig. 2c is shown the result of the feature vector with a 2D stick figure model. The model is the simplest way to represent a human body, and thus it is relative easier to fit into the link-parts from the anthropological information. Therefore, we construct a skeleton for the stick figure body that is including five main body parts: a torso, two arms, and two legs. The body parts are defined: a torso, two upper arms, two forearms, two thighs, two shins, and two feet. Each arm is a joint of an upper arm and a forearm at the elbow. Each leg is a joint of a thing and a shin at the knee as shown in Fig. 3a and Fig. 3b. For the association of the eleven body parts, we introduce fifteen joints: three joints for each arm (a shoulder, an elbow, and a wrist), four joints for each leg (a hip, an ankle and a foot), and one joint for torso as shown in Fig. 3b and Fig. 3c. The skeleton feature vector κ is represented as:

$$\kappa = \left\{k_1^x, k_1^y, k_2^x, k_2^y, \ldots, k_{|R|}^x, k_{|R|}^y\right\} \quad (1)$$

where $k_{|R|}^x$ and $k_{|R|}^y$ are x and y elements of the normalized 2D vector with respect to ith joints, and $|R|$ is the number of joints in model. The 2D stick figure model consists of body parts that are connected by joints which are called segmental reference points, $R = \{r_1, \ldots, r_i, \ldots, r_{|R|}\}$. The v joint angles that are defined as d, $d = \{\theta_1, \ldots \theta_i, \ldots, \theta_{|v|}\}$, where $|v|$ is the number of joint angles, i is defined the i th reference position at i (r_i). The θ_i-angles are defined as the joint angles at r_i which are placed in a sagittal plane on an x-axis, $\hat{e}_1^{(i)}$ and y-axis, $\hat{e}_2^{(i)}$ as shown in Fig. 2d. Next step, we analyze the set of skeleton features with body energy action model. The 2D stick figure sample images are shown in Fig. 4.

III. Body Energy Action Model

Human action interpreted through the Body Energy Action Model which is adapted from a fundamental principle of the Biomechanics [24]. The idea is that the regular physical activity is usually perform on the days of the week, called basic daily posture which takes a moderate energy cost such as sitting, standing, walking. The physical inactivity is rarely showing the posture that has the higher energy cost than the energy of basic daily posture. For example, the arm moves up in 180° in standing posture that take the energy more than normally standing posture. Therefore, we incorporate these concepts into the body energy action model. The model has captured the essence of action that the energy expenditure of entire body computed at its joints is such a measure. It is considered the discrete action classification process into two stages. In the first stage, we analyze the entire body with energy expenditure into primitive actions, including lying down, sitting, and standing. The lying has less energy consumption than other classes and jumping has the highest energy. Next, we computed the force at each joint in body parts. The energy can be classified into six classes: Standing, running, jumping, bending, sitting and lying.

A. The Energy Analysis in Whole-Body

The first step of analysis is finding the energy in whole body. We used κ that is corresponding to input action from stick figure. We applied the gravity forces that related from entire body. Let ψ be the energy of entire body based on the anatomical data. Data is an average percentage of a segmental body part, including the segmental length ς_i, and the mass m_i as shown in Appendix A. The ψ is calculated from the weight average of the total body segments that related to gravity forces equal to 9.8 meters/square sec. We can express the global energy by

$$\psi = \left(\sum_{i=1}^{|R|} m_i \cdot g \cdot h_i\right) / \left(\sum_{i=1}^{|R|} m_i \cdot g\right) \quad (2)$$

where

$$h_i = \sum_{j=1}^{|R|} \varsigma_j \cdot \delta_j \quad (3)$$

The h_i is the segmental height from the segmental length ζ_i. All other points on the body parts are found by

the rotation matrix δ_i with θ_i degree that aligns on the $\hat{e}_1^{(i)}$ and $\hat{e}_2^{(i)}$ as shown in Fig. 2d.

$$\delta_i = \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix} \cdot \begin{bmatrix} \hat{e}_1^{(i)} \\ \hat{e}_2^{(i)} \end{bmatrix} \quad (4)$$

where

$$\hat{e}_1^{(i)} = \cos\theta_i \cdot e_1^{(i-1)} + \sin\theta_i \cdot e_2^{(i-1)} \quad (5)$$

$$\hat{e}_2^{(i)} = \cos\theta_i \cdot e_2^{(i-1)} - \sin\theta_i \cdot e_1^{(i-1)} \quad (6)$$

We added the potential energy. It can be measured the variety action meaning by using each part for representing the more detail of actions. Potential energy is energy due to its of which position related to the object's weight and its elevation or height above the ground. It would be more useful to have energy to describe the force in every joint. Let $\beta = \{\partial_1,..,\partial_{|R|}\}$ be a set of the potential energies depending on joint positions and their corresponding gravity forces. The ∂_i is defined as follows:

$$\partial_i = \varsigma_i \cdot h_i. \quad (7)$$

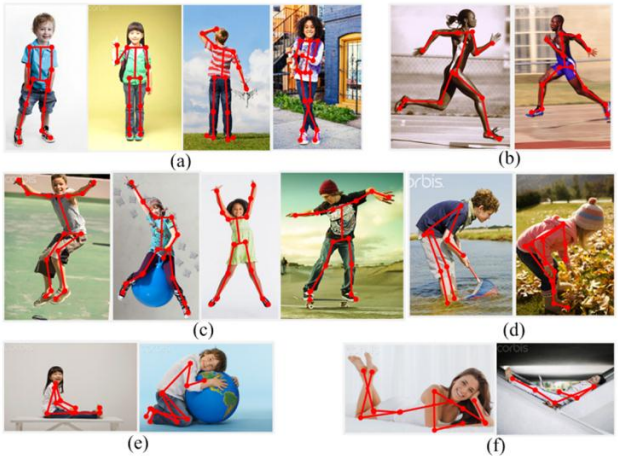

Figure 4. Sample images of the still image action dataset. The location of 15 joints have been evaluated on each image. (a) standing (b) running (c) jumping (d) bending (e) sitting (f) lying..

B. The Energy Analysis in Body Parts

In the second stage, we emphasized to find the energy intensity of each part that acts on segmental joints for classifying primitive actions. The energy used in this stage called energy intensity that can be evaluated from the resorting force to slight perturbations imposed in static positions. The joint torque produced by muscles to perform a task consists of two components, the torque to compensate the external force F_i^e from environment force and the torque τ_i^b necessary to move the body part. The standard formulation for the equation of torque is,

$$\tau_i = F_i^e + \tau_i^b \quad (8)$$

Since the force extraction of body movement is concerned, thus we assume that τ_i is the energy generated by movements of limbs at joints and the entire body. A turning effect would have been produced by the position and orientation of the line of action of the force as well as by its size. Let $\Gamma = \{\tau_1,..,\tau_i,..,\tau_{|R|}\}$ be a set of torques that relates the force size F_i, and the distance λ_i at joint r_i. The τ_i^b is mathematically defined as follows:

$$\tau_i^b = F_i \times \lambda_i \quad (9)$$

The F_i corresponds to m_i and g. The equation of F_i is the forces at joint r_i that corresponding to weight.

$$F_i = \sum_i m_i.g \quad (10)$$

where m is the mass at r_i, the g is the gravitational constant. The λ_i is the distance between the i^{th} reference position at i, r_i and the base reference point r_0. The equation of distance λ_i is

$$\lambda_i = \sum_{j=1}^{j=i} P^{(j,j-1)} \quad (11)$$

where $P^{(i,j-1)}$ is estimated from the rotation matrix that aligns on $\hat{e}_1^{(i)}$, and $\hat{e}_2^{(i)}$.

$$P^{(i,j-1)} = l_i \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix} \cdot \begin{bmatrix} \hat{e}_1^{(i-1)} \\ \hat{e}_2^{(i-1)} \end{bmatrix}.$$

Therefore, the body action energy of each part is formulated as the summation of transform matrices of all body parts.

IV. Sematic Classification

The goal of semantic classification is to arrange skeletons according to the body's energy intensity into C categories, $\{C_1, C_2..., C_c\}$. We compare the results with traditional five classifiers: Naive-Bayes, the multi-layer perception networks (MLPN), Support Vector Machine (SVM), and Kernel function. Next, we descript the two classifiers: SVM and Kernel.

A. Support Vector Machines

Support Vector Machine (SVM) [26] is another simple classifier that is widely use in CBIR [27]. SVM is formalized as an optimization problem which finds the best hyperplane separating relevant and irrelevant which belong to in two different classes. Let $(x_i, y_i)_{1 \le i \le N}$ be a set of training examples, each example $x_i \in \mathbb{R}^d$, d being the dimensional of input feature space, belongs to a class labeled by $y_i \in \{-1, 1\}$. We used is a Gaussian radial basis function, the corresponding feature space is a Hilbert space of infinite dimension. Maximum margin classifiers are well regularized and the infinite dimension does not

spoil the results. In a two-class case, the decision function for a test sample x has the following form:

$$g(x) = \sum_i \alpha_i y_i K(x_i, x) - b \text{ when } \propto_i > 0 \quad (12)$$

where x is the test sample, $\propto_i$ the learned weight of the training sample x_i, and b is learned threshold parameter. The training instances that lie closest to the hyperplane in the transformed space are called support vectors. The number of these support vectors is usually small compared to the size of the training set and they determine the margin of the hyperplane, and thus the decision surface.

B. Kernel Function

During the SVM model generation, the input vectors, are mapped into a new higher dimensional feature space. Then, an optimal separating hyperplane in the new feature space is constructed by a kernel function which products between input vectors x and $K(x, y)$. Two most used kernel functions are Polynomial and Gaussian Radial Basis Function (RBF) kernel functions which are:

$$K_{poly}(x_i, y_i) = (x_i \cdot y_i + 1)^p, \quad (13)$$

where p is the degree of polynomial.

and $$K_{gaussian}(x_i, y_i) = e^{-\|x_i - y_i\|^2 / 2\sigma^2} \quad (14)$$

where $\|\ldots\|$ denotes the L_2 norm, x and y are two sample vectors, and σ is the width of the Gaussian kernel, generally determined using cross-validation. All vectors lying on one side of the hyperplane are labelled as -1, and all vectors lying on another side are labeled as +1. RBF kernels have exhibited good generalization properties in many classification problems. However, the use of a simple Euclidian distance implies small variations on the kernel value in high dimensional feature spaces.

C. Kernel Based Data Fusion

Kernel fusion is a popular scheme to use a classifier to learn the relations between modality components at different abstraction levels. Merging all the descriptors into a single flat classifier leads to a fully integrated fusion strategy since the fusion classifier obtains all the information from all sources. Consequently, when using a RBF kernel, a single σ parameter is expected to "fit" properly the sample vectors relations, whereas it makes much more sense to train a combined RBF kernel using one σ per modality. Combination of unimodal kernels leads to keep as much information as possible from each modality. Therefore, RBF kernel is integrated into the combine kernel has the following form:

$$K_{combine}(x, y) = C.K_j(x_j, y_j) \quad (15)$$

when $(1 \leq j \leq |j|)$

where x and y is sample value. C is the combining function over the |j| modalities, x_j and y_j are sample vectors for modality j. This kernel scheme is learning the regularities formed by the components independently from the modalities. It is easy to use as it just consists in concatenating the various data in a single vector.

V. Experimental Results

The purpose of the human action classification is recognized more authentic intuitive meaning of actions to the visual perception. We examined the action classification into the six general action classes including standing, running, jumping, bending, sitting, and lying.

A. Dataset and Image Categorization

In our experiments, we manually selected the probe images from data sources. The probe images are manually selected and annotated from LabelMe. Dataset is contained approximately 1,500 images. We setup a dataset to cover a variety of contents including background, foreground objects, and dominant human. The background and foreground was scoped into indoor (restaurant, store, office, home) and the outdoors (city, park, beach, street). For human, we considered the human centric images where all parts of the human are visually clear for identifying.

B. Evaluation Methods

In this section, we evaluated the semantic classification results by comparing with five methods [26]-[28]; Naive-Bayes, the multi-layer perception networks (MLPN), Support Vector Machine (SVM), and Kernel based Data Fusion. We used open source software Waikato Environment for Knowledge Analysis (WEKA) [29] to apply a traditional classification. We have been applied precision, recall, F-measure, and accuracy. Precision is defined to the total numbers of retrieved images with all corpuses, while recall is the specific related image with retrieval images. The highest value of the both measurements is 1. Their definitions are shown below.

$$\text{precision}_i = \frac{\#\text{ of correctly classified images of class } i}{\#\text{of images classified to class } i}$$

$$\text{recall}_i = \frac{\#\text{ of correctly classified images of class } i}{\#\text{of images in the class } i}$$

$$F\text{-}measure_i = \frac{2 \cdot \text{precision}_i \cdot \text{recall}_i}{\text{precision}_i + \text{recall}_i}$$

$$\text{accuracy} = \frac{\#\text{ of correctly classified images}}{\#\text{of images}}$$

C. Experimental Results

The experiment, we compare five types classification method for testing the body energy action model. We use all the features consisting of set of forces (ψ, F, Γ) from 15 reference points. Table I to V show the confusion matrices and evaluation methods. Each column of the matrix represents one class and shows how the instances of this class are classified. Each row represents the instances that are predicted to belong to a given class, and shows the true classes of these instances.

TABLE I. CONFUSION MATRIX OF THE CLASSIFICATION RESULTS WITH NAÏVE-BAYES

	walking	running	bending	jumping	sitting	lying	Performance(%) Pr	Recall	F1
walking	0.73	0.16	0.08	0.06	0.03	0	71.6	68.9	70.2
running	0.14	0.7	0.11	0.02	0.05	0	71.4	68.6	70.0
bending	0.01	0.03	0.82	0.01	0.12	0	78.1	82.8	80.4
jumping	0.07	0.05	0.01	0.67	0.15	0.09	69.8	64.4	67.0
sitting	0.07	0.04	0.01	0.12	0.75	0.13	58.1	67.0	62.2
lying	0	0	0.02	0.08	0.19	0.79	78.2	73.1	75.6
Average							70.7		

TABLE II. CONFUSION MATRIX OF THE CLASSIFICATION RESULTS WITH THE MULTI-LAYER PERCEPTION NETWORKS

	walking	running	bending	jumping	sitting	lying	Performance(%) Pr	Recall	F1
walking	0.76	0.14	0.08	0.04	0.02	0	75.2	73.1	74.1
running	0.08	0.78	0.09	0.01	0.05	0	72.9	77.2	75.0
bending	0.05	0.02	0.81	0.01	0.02	0.02	78.6	87.1	82.7
jumping	0.04	0.07	0.01	0.69	0.14	0.05	75.8	69.0	72.3
sitting	0.08	0.06	0.04	0.09	0.72	0.06	70.6	68.6	69.6
lying	0	0	0	0.07	0.07	0.87	87.0	86.1	86.6
Average							76.7		

TABLE III. CONFUSION MATRIX OF THE CLASSIFICATION RESULTS WITH SOM

	walking	running	bending	jumping	sitting	lying	Performance(%) Pr	Recall	F1
walking	0.83	0.09	0.05	0.04	0.03	0	77.6	79.8	78.7
running	0.09	0.75	0.11	0.02	0.05	0	75.0	73.5	74.3
bending	0.07	0.09	0.82	0.01	0.02	0	82.0	81.2	81.6
jumping	0.05	0.03	0.01	0.71	0.15	0.06	74.0	70.3	72.1
sitting	0.03	0.04	0.01	0.11	0.81	0.07	72.3	75.7	74.0
lying	0	0	0	0.07	0.06	0.94	87.9	87.9	87.9
Average							78.1		

TABLE IV. CONFUSION MATRIX OF THE CLASSIFICATION RESULTS WITH SVM

	walking	running	bending	jumping	sitting	lying	Performance(%) Pr	Recall	F1
walking	0.84	0.11	0.09	0.01	0.01	0.00	81.6	79.2	80.4
running	0.09	0.82	0.05	0.02	0.02	0.00	81.2	82.0	81.6
bending	0.07	0.05	0.83	0.01	0.00	0.00	84.7	86.5	85.6
jumping	0.03	0.02	0.01	0.78	0.09	0.05	83.9	79.6	81.7
sitting	0.00	0.01	0.00	0.06	0.89	0.03	84.0	89.9	86.8
lying	0.00	0.00	0.00	0.05	0.05	0.95	92.2	90.5	91.3
Average							84.6		

TABLE V. CONFUSION MATRIX OF THE CLASSIFICATION RESULTS WITH KERNEL BASE DATA FUSION

	walking	running	bending	jumping	sitting	lying	Performance(%) Pr	Recall	F1
walking	0.83	0.04	0.02	0.02	0.01	0	93.3	90.2	91.7
running	0.03	0.82	0.05	0.02	0.02	0	85.4	87.2	86.3
bending	0.02	0.05	0.88	0	0	0	91.7	92.6	92.1
jumping	0.01	0.02	0.01	0.87	0.09	0.03	87.9	84.5	86.1
sitting	0	0.03	0	0.06	0.91	0.02	84.3	89.2	86.7
lying	0	0	0	0.02	0.05	0.95	95.0	93.1	94.1
Average							89.5		

The performance of the method was evaluated using 5-fold cross validation method with $p-$values . The smallest $p-$values reflect the most discriminative features. Comparing the results, we can observe that the accuracy of walking class in SVM provide higher than other methods as shown in Table IV. The results of lying class in kernel function and SVM can achieve the better accuracy of 95% when the lying class in Naïve-Bayes gains 79%. We can see that lying and bending postures are easily to classify with whole energy from main body. Whereas the jumping and running class with the Naïve-Bayes only 67% and 70%, kernel function up to 87% as show in Table I. The kernel base data fusion seems the best classifier since it can produce the highest average accuracy of 89.5%, compared to 84.6% for SVM, 78.1% for SOM, 76.7% for MLPN, and 70.7% for Naïve-Bayes.

We are concluding that, the accuracy of the kernel function based on body action energy model that are suitable for semantic classification. The results of kernel function shown in Table V. The example of classification results as shown in Fig. 5.

Figure 5. Sample images of the still image action dataset. (a) Standing (b) Running (c) Jumping (d) Bending

VI. CONCLUSION

Researchers have attempted available methodologies and techniques to interpret the semantic images. This paper proposed a novel technique for classifying the semantic images. The major feature components consist of the set of body energy from reference points. The energy features can be computing form the 2D stick figure model. Then, the images can be classified into the

human action. The results indicated that the proposed method offers good classification of semantics. To improve the algorithm to be able to classify more the group of semantic concepts and human activities by adding more human features is interesting for the future work.

REFERENCES

[1] T. Zhang, J. Liu, S. Liu, C. Xu, S. Member, and H. Lu, "Boosted exemplar learning for action recognition and annotation," *IEEE Trans. on Circuit and Systems for Video Technology*, vol. 21, no. 7, pp. 853–866, July 2011.

[2] T. Chen, P. Tan, L. Q. Ma, M. M. Cheng, A. Shamir, and S. M. Hu, "PoseShop: Human image database construction and personalized content synthesis," *IEEE Trans. on Visualization and Computer Graphics*, vol. 19, no. 5, pp. 824–37, May 2013.

[3] L. Cao, J. Luo, H. Kautz, T. S. Huang, and L. Fellow, "Image annotation within the context of personal photo collections using hierarchical event and scene models," *IEEE Trans. on Multimedia*, vol. 11, no. 2, pp. 208–219, Feb. 2009.

[4] I. Classification, "Semantic personal image classification by energy expenditure," presented at the ISCIT, pp. 1072–1075, Oct. 12-14, 2005.

[5] E. Molitorisová, "Automatic semantic image annotation system," presented at the PHDCONF, pp. 414–419, 2012.

[6] M. Wang, X. Zhou, and T.-S. Chua, "Automatic image annotation via local multi-label classification," in *Proc. 2008 Int. Conf. on Content-based Image and Video Retrieval*, 2008, pp. 17-26.

[7] J. Yoo and M. S. Nixon, "Automated markerless analysis of human gait motion for recognition and classification," *ETRI Journal*, pp. 259–266, April 2011.

[8] J. Liu, B. Kuipers, and S. Savarese, "Recognizing human actions by attributes," presented at the IEEE Conf. on CVPR, pp. 3337–3344, June 2011.

[9] D. Kuettel, M. Guillaumin, and V. Ferrari, "Combining image-level and segment-level models for automatic annotation," *Advances in Multimedia Modeling Lecture Notes in Computer Science*, vol. 7131, pp. 17–28, 2012.

[10] J. Fan, Y. Gao, and H. Luo, "Hierarchical classification for automatic image annotation," in *Proc. 30th Annu. Int. ACM SIGIR Conf. on Research and Development in Inf. Retrieval*, 2007, pp. 111–118.

[11] S. Dasiopoulou and E. Giannakidou, "A survey of semantic image and video annotation tools," *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution Lecture Notes in Computer Science*, vol. 6050, pp. 196–239, 2011.

[12] H. Bouyerbou and S. Oukid, "Hybrid image representation methods for automatic image annotation: A survey," presented at the ICSES, 2012.

[13] B. A. Torralba, B. C. Russell, J. Yuen, and A. Torralba, "LabelMe : Online image annotation and applications," in *Proc. IEEE*, vol. 98, 2010, pp. 1467–1484.

[14] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe : A database and web-based tool for image annotation," *Int. Journal of Computer Vision*, vol. 77, no. 1–3, pp. 157–173, 2008.

[15] J. Álvez, J. Atserias, J. Carrera, S. Climent, E. Laparra, and G. Rigau, "Complete and consistent annotation of WordNet using the top concept ontology," in *Proc. 6th Int. Conf. on Language Resources and Evaluation*, 1998, pp. 1529-1534.

[16] A. Sanfilippo and S. Tratz, "Ontological annotation with WordNet," presented at the 5th Int. Workshop on Knowledge Markup and Semantic Annotation, pp. 27–36, 2006.

[17] H. Langone, B. Haskell, and G. Miller, "Annotating WordNet," in *Proc. Workshop Frontiers in Corpus Annotation*, 2004, pp. 63–69.

[18] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas, and L. Fei-Fei, "Human action recognition by learning bases of action attributes and parts," presented at the Int. Conf. on Computer Vision, pp. 1331–1338, Nov. 2011.

[19] A. Shabani, J. Zelek, and D. Clausi, "Human action recognition using salient opponent-based motion features," presented at the 7th Canadian Conf. on Computer and Robotic Vision, Ottawa, pp. 362 – 369, 2010.

[20] S. Yoon and A. Kuijper, "Human action recognition using segmented skeletal features," presented at the 20th ICPR, 2010.

[21] N. Ikizler and R. Cinbis, "Recognizing actions from still images," presented at the 19th ICPR, Dec. 8-11, 2008.

[22] G. Sukthankar and K. Sycara, "A cost minimization approach to human behavior Recognition," in *Proc. 4th Int. Joint Conf. on Autonomous Agents and Multi-Agent Systems*, 2005.

[23] R. Arnheim, *Art and Visual Perception A Psychology of the Creative Eye*, University of California Press, 1974, pp. 11–15.

[24] D. A. Winter, *Biomechanics and Motor Control of Human Movement*, 3rd ed. John Wiley & Sons, 2005.

[25] J. Park and H. Oh, "Human posture recognition using curved segments for image retrieval," in *Proc. Storage and Retrieval for Media Databases 2000*, 1999, vol. 3972.

[26] D. G. S. R. O. Duda and P. E. Hart, *Pattern Classification*, 2nd ed. New York Wiley, 2001.

[27] Y. Luo, M. Liao, J. Yan, and C. Zhang, "A multi-features fusion support vector machine method (MF-SVM) for classification of mangrove remote sensing image," *Journal of Computational Information Systems*, vol. 1, pp. 323–334, Jan. 2012.

[28] T. Mitchell, *Machine Learning*, McGraw Hill, 1997.

[29] Waikato Environment for Knowledge Analysis (WEKA). [Online]. Available: http:// www.cs.waikato.ac.nz/ml/weka.

Nutchanun Chinpanthana received the B.S. degree in computer science and was awarded the university for the best graduating student from the University of the Thai Chamber of Commerce in 1994, M.S. in Applied Statistics and Information Technology from National Institute of Development Administration in 1997. She is currently a lecturer in Faculty of Information Technology, Dhurakij Pundit University, Thailand. Her research interests include content-based image retrieval, image processing, computer vision, and semantic images.

Tejtasin Phiasai received the B.Eng in Telecommunication Engineering from King Mongkut's Institute of Technology Ladkrabang (KMITL) and M.Eng in Electrical Engineering from King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand, in 1999, 2001, respectively. He is currently working toward the Ph.D. degree in Electrical and Computer Engineering of KMUTT. His research interests include signal processing and computer vision.

