International Journal of Information System Modeling and Design
Volume 13 • Issue 2

Enterprise Management Data Acquisition System Based on WoT

Kun Yan, Jilin Normal University, China*
Dongyan Li, Jilin Normal University, China

ABSTRACT

With the rapid development of the WoT, acquisition requirements to the data grow with each passing day. However, the existing data acquisition systems usually use the upload mode. Uploading data not only takes a lot of time, but also requires staff to continuously monitor on duty. Aiming at the problem of enterprise management data acquisition, a WoT-based enterprise management data acquisition and analysis system is constructed. The system obtains multiple monitoring data through the data acquisition module and integrates these data through the processing module, which can solve the problems of traditional data acquisition systems. In order to effectively fuse data, based on canonical correlation analysis (CCA), the multiple order statistics correlation discrimination analysis algorithm is proposed as a fusion strategy. Experimental results show that the fusion strategy based on CCA can effectively integrate monitoring data and accurately identify different characters.

KEYWORDS

Canonical Correlation Analysis, Data Acquisition System, Monitoring Data, Web of Things

1. INTRODUCTION

The Web of Things (WoT) refers to real-time acquisition of any objects or processes that need to be monitored, connected, and interacted through various information sensors. Specifically, it collects various required information such as sound, light, heat, electricity, mechanics, chemistry, biology, location, etc., and realizes the ubiquitous connection between things and things, things and people, and realizes the intelligent perception, recognition and management of things and processes. The WoT is an information carrier based on the internet and traditional telecommunications networks, which allows all ordinary physical objects that can be independently addressed to form an interconnected network (Liu 2011). In 1999, the Massachusetts Institute of Technology established the "Automatic Identification Center" and proposed that "everything can be connected through the network", which clarified the basic meaning of WoT (Liu 2012). Compared with the traditional internet, WoT has its own distinctive features. First of all, it is a wide application of various perception technologies. A large number of sensors of various types are deployed on the WoT, each sensor is an information source, and different types of sensors capture information in different formats. The data obtained by the sensor is of real-time nature, collecting environmental information periodically at a certain frequency and continuously updating the data. Secondly, the WoT not only provides the connection

DOI: 10.4018/IJISMD.313577 *Corresponding Author

of sensors, it also has the ability of intelligent processing, which can implement intelligent control of objects. The WoT combines sensors and intelligent processing, and uses various intelligent technologies such as cloud computing and pattern recognition to expand application areas. WoT analyzes and processes meaningful data from the massive information obtained from sensors to adapt to different needs and discover new application fields and application modes. In the world of WoT, all sensing devices are connected to form a multi-tactile sensing network. This requires our sensing devices not only to "feel" but also to "recognize" and "analyze", and transmit the information after identification and analysis to the back, the end uses the center to prepare for decision-making. And video data has become the key target of intelligent recognition due to its unique analytical and image processing technology accumulation.

The data acquisition system is an important part of WoT, which can provide real-time data for management and control, and provide an important basis for equipment monitoring and performance analysis for operators. In order to shorten the system development cycle and avoid the repeated design of the system, the design and research of the data acquisition system is of great significance. At present, the data that needs to be collected mainly comes from various sensors, which are characterized by wide spatial distribution and large amount of data. In some high-risk and high-intensity special environments, it is impossible to obtain data manually. At this time, it is necessary to obtain the temperature, humidity, pressure and other data of the work site through the sensor, and to remotely transmit and process the collected data through the wireless network. The research on the data acquisition system first started in the United States. In 1956, the United States studied a set of military test systems, aims to test the relevance of the system files. And the test tasks were completed by the automatic control of the equipment. Because this kind of acquisition system has fast data processing speed, high flexibility, and can meet data acquisition tasks that cannot be completed by traditional methods, it has been initially recognized. In the 1960s, a complete set of acquisition equipment with complete functions was introduced into the foreign market, and most of these equipment were dedicated systems. In the 1970s, the combination of computer technology and data acquisition equipment gave birth to a new generation of data acquisition systems. And this type of system has achieved rapid development by virtue of its excellent performance. In the 1980s, with the promotion and popularization of computers, general-purpose acquisition systems and automatic test systems came into being, which further promoted the development of data acquisition systems. Since the 1990s, in countries with advanced technology in the world, the data acquisition technology has been widely used in military, avionics and aerospace technology, industry and other fields (Chen 2012).

The WoT requires the deployment of a large number of sensor nodes to obtain environmental information, such that the sensor data usually having the characteristics of massiveness, heterogeneity, uncertainty and relevance (Li 2011, Li 2009). Because the sensor nodes use broadcast for wireless communication, a large amount of data is transmitted in the network, which inevitably brings about network congestion, data loss and delay problems. Take the surveillance video system as an example, the traditional surveillance system collects surveillance video images through front-end cameras, and then uploads them to the back-end server for storage and processing, and obtains useful information in surveillance video images through human inspection or some related algorithms. This traditional monitoring system has two shortcomings. One is that it requires a lot of storage space. The other is that the monitoring system is sensitive to time. It requires sensor nodes to transmit the collected data to the central processing node in real time, and outdated information is not only useless to the system, but may also bring catastrophic consequences. In addition, the transmission of a large amount of redundant information accelerates the loss of battery energy of the sensor node, and seriously affects the service life of the network. Therefore, the cleaning and fusion processing of the collected data before storage is indispensable. Specifically, it is necessary not only to ensure the real-time processing of these changeable and unknown data, but also to integrate the data from multiple collection nodes, to fuse and converge to the central processing node according to a specific algorithm, so as to meet the high-level application system's demand for reliable and efficient data collection.

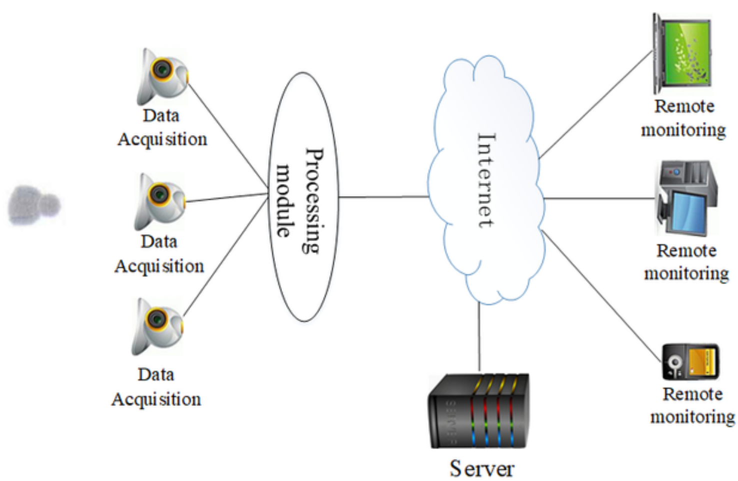
According to the above problems, this paper is oriented to the enterprise management data acquisition system, taking the video face recognition technology in the surveillance system as the entry point, and constructing a WoT-based enterprise management data acquisition and analysis system. This system obtains multiple monitoring data through the data acquisition module, and conducts fusion analysis on the data through the processing module, which solves the problems of traditional data acquisition systems such as inability to respond in real time, occupied storage and excessive data transmission. The following part of this paper is organized as follows: the architecture of WoT-based enterprise management data acquisition and analysis system is proposed in Section 2; the multiple order statistics correlation discrimination analysis algorithm is used as the fusion strategy for surveillance video data in Section 3; the experiments are provided in Section 4; Section 5 is the conclusion and discussion.

2. ARCHITECTURE OF WOT-BASED ENTERPRISE MANAGEMENT DATA ACQUISITION AND ANALYSIS SYSTEM

In order to solve the problems of traditional data acquisition systems that cannot respond in real time, occupy storage and data transmission volume is too large, etc., an architecture based on WoT is proposed for enterprise management data acquisition system, which is shown in Figure 1.

In Figure 1, according to the function, the system can be divided into three parts, namely the perception layer, the transmission layer, and the application layer. The perception layer is composed of a data acquisition module (web camera) and a processing module. The camera collects the image and transmits it to the middleware server through the network after the preliminary processing (data fusion) of the processing module. The server performs operations such as data storage, identification, and early warning. The transmission layer obtains the middleware server data through programs and displays the processing results on the application layer.

Figure 1. The illustration of the architecture for WoT-based enterprise management data acquisition and analysis system.



Let us suppose that the data acquisition module has collected two surveillance video sequences of someone: $S_1 = \left\{ s_1^1, \ldots, s_{n_1}^1 \right\}$ and $S_2 = \left\{ s_1^2, \ldots, s_{n_2}^2 \right\}$, where s_j^i denotes the j^{th} frame in sequence i, n_1 and n_2 denotes the number of frames in each video. The data processing module performs data fusion on the video sequences S_1 and S_2. Specifically, we first use the first-order statistical features and second-order statistical features to model S_1 and S_2 respectively.

First-order statistics: Computing the mean vector m_1 of video sequence S_1, it shows the average position of the video in the Euclidean space:

$$m_1 = \frac{1}{n_1}\sum\nolimits_{j=1}^{n_1} s_j^1. \tag{1}$$

Second-order statistics: The second-order statistics feature, i.e., covariance matrix C_2 of video sequence S_2 is computed as follows, which represent the correlation of two individual features of each pair of samples in the video:

$$C_2 = \frac{1}{n_2 - 1}\sum\nolimits_{j=1}^{n_2} \left(s_j^2 - m_2\right)(s_j^2 - m_2)^T. \tag{2}$$

where m_2 is the mean vector of video S_2.

Then, in order to overcome the heterogeneity between modeling representations m_1 and C_2, different kernel functions are used to map them to high-dimensional Euclidean space. Finally, in the high-dimensional space, low-dimensional embeddings W_m and W_C are used to map them to the low-dimensional shared subspace for data fusion.

3. MULTIPLE ORDER STATISTICS CORRELATION DISCRIMINATION ANALYSIS

In this section, based on canonical correlation analysis (CCA), the multiple order statistics correlation discrimination analysis (MOSCDA) algorithm is developed to fuse surveillance video data. Suppose that we have two video monitors, and have collected $2N$ video sequences: $\mathbb{S}^1 = \left\{S_1^1, S_2^1, \cdots, S_N^1\right\}$ and $\mathbb{S}^2 = \left\{S_1^2, S_2^2, \cdots, S_N^2\right\}$, that is each person has two videos at each time. In this paper, we use the first-order statistics to represent the first video and use the second-order statistics features to represent the second video.

Different order statistics information characterizes the video from different perspectives. For instance, the mean vector roughly reflects the position of the sample in the Euclidean space, the covariance matrix denotes the variance and the correlation of features in the diagonal elements and the non-diagonal elements, respectively. Hence, these statistics features can provide complementary information to represent the same person.

From above, we know that each person (has two videos) can be modeled by a variable pair $\left(m, C\right)$, this is a classical two-model video classification problem. It also can be viewed as multi-view features of an instance. Therefore, the common multi-view learning method CCA can be used to extract and fuse the statistics features.

However, the first-order statistics features lie on the Euclidean space, while the second-order statistics features lie on Riemannian manifold (SPD manifold) $\mathbb{S}_{++}^d$. It is difficult for CCA to project them into the common subspace directly. To reduce the dissimilarity between the heterogeneous spaces, we first embed the Riemannian manifold $\mathbb{S}_{++}^d$ into a high dimensional Euclidean space by using the Logarithm kernel:

$$k_{log}\left(C_i, C_j\right) = tr\left[log\left(C_i\right) \cdot log\left(C_j\right)\right]. \tag{3}$$

After the embedding, CCA based method is learned from the high dimensional Euclidean space and the original Euclidean space (or its corresponding Hilbert space) to the final common subspace.

From above we know that different order statistics features contain different discriminative information, how to fuse them is an important problem.

Canonical correlation analysis is a famous multi-view learning method, it seeks a pair of projection vectors such that the correlation is maximized. Suppose $\mathbb{S}^1 = \left\{S_1^1, S_2^1, \cdots, S_N^1\right\}$ and $\mathbb{S}^2 = \left\{S_1^2, S_2^2, \cdots, S_N^2\right\}$ denotes $2N$ videos. $m = \left\{m_1, m_2, \cdots, m_N\right\}$ and $\mathbb{C} = \left\{C_1, C_2, \cdots, C_N\right\}$ are the videos of extracted first-order statistics and second-order statistics features respectively, where $m_i \in \mathbb{R}^d, C_i \in \mathbb{R}^{d\times d}, i = 1, \cdots, N$.

Since the two variable groups m and $\mathbb{C}$ lie on heterogeneous spaces, the implicit mapping function $\varphi : \mathbb{R}^d \to \mathcal{H}$ and $\phi : \mathcal{M} \to \mathcal{T}$ are used simultaneously to map m and $\mathbb{C}$ to the high-dimensional Euclidean spaces. In the high-dimensional Euclidean spaces, kernel CCA (KCCA) seeks the two transforms w_m and W_C, to project $\varphi\left(m\right)$ and $\phi\left(C\right)$ to the common subspace. In other words, the correlation between variables $< w_m, \varphi\left(m\right) >_v$ and $< W_c, \phi\left(C\right) >_{M_t}$ is maximized, that is

$$\max_{w_m, W_c} \quad \frac{cov(< w_m, \varphi\left(m\right) >_v, < W_c, \phi\left(C\right) >_{M_t})}{\sqrt{var(< w_m, \varphi\left(m\right) >_v) var(< W_c, \phi\left(C\right) >_{M_t})}} \tag{4}$$

where $< \; >_v$ and $< \; >_{M_t}$ tare vector inner product and matrix inner product, respectively.

Let $w_m = \sum_{i=1}^N \alpha_i \varphi\left(m_i\right)$, $W_c = \sum_{i=1}^N \beta_i \phi\left(C_i\right)$, then $< w_m, \varphi\left(m\right) >_v = \alpha^T K\left(m, m\right)$, $< W_c, \phi\left(C\right) >_{M_t} = \beta^T K_{log}\left(\mathbb{C}, C\right)$, where

$$\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}, \; K\left(m, m\right) = \begin{bmatrix} k\left(m_1, m\right) \\ \vdots \\ k\left(m_N, m\right) \end{bmatrix}, \; \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}, \; K_{log}\left(\mathbb{C}, C\right) = \begin{bmatrix} k_{log}\left(C_1, C\right) \\ \vdots \\ k_{log}\left(C_N, C\right) \end{bmatrix}. \tag{5}$$

The problem (4) can be rewritten as:

$$\begin{aligned} \arg\max_{\alpha,\, \beta} \quad & \alpha^T G_{mc} \; \beta \\ s.t. \quad & \alpha^T G_m \; \alpha = 1, \\ & \beta^T G_c \; \beta = 1, \end{aligned} \tag{6}$$

where

$$
\begin{aligned}
G_{mc} &= \frac{1}{N-1}\sum_{i=1}^{N}\tilde{K}(\mathrm{m}, m_i)(\widetilde{K_{log}}(\mathbb{C}, C_i))^T, \\
G_{m} &= \frac{1}{N-1}\sum_{i=1}^{N}\tilde{K}(\mathrm{m}, m_i)(\tilde{K}(\mathrm{m}, m_i))^T, \\
G_{c} &= \frac{1}{N-1}\sum_{i=1}^{N}\widetilde{K_{log}}(\mathbb{C}, C_i)(\widetilde{K_{log}}(\mathbb{C}, C_i))^T,
\end{aligned}
\tag{7}
$$

and here $\tilde{K}\left(m, m_i\right) = K\left(m, m_i\right) - \bar{K}$, $\widetilde{K_{log}}\left(\mathbb{C}, C_i\right) = K_{log}\left(\mathbb{C}, C_i\right) - \overline{K_{log}}$, where $\bar{K} = \frac{1}{N}\sum_{i=1}^{N} K\left(m, m_i\right)$, $\overline{K_{log}} = \frac{1}{N}\sum_{i=1}^{N} K_{log}\left(\mathbb{C}, C_i\right)$.

Using the Lagrange multiplier method, the following two generalized eigenvalue equations are obtained:

$$
\begin{aligned}
(G_m)^{-1} G_{mc} (G_c)^{-1} G_{cm} \alpha &= \lambda^2 \alpha, \\
(G_c)^{-1} G_{cm} (G_m)^{-1} G_{mc} \beta &= \lambda^2 \beta,
\end{aligned}
\tag{8}
$$

where $G_{cm} = G_{mc}^T$.

Solving the eigen-problem (8), we can get the p largest eigenvectors $A = \left[\alpha_1, \cdots, \alpha_p\right]$ and $B = \left[\beta_1, \cdots, \beta_p\right]$. Given a pair of examples $\left(m_t, C_t\right)$, the projections in the discriminant subspace are obtained by

$$
A^T K\left(m, m_t\right) \ and \ B^T K_{log}\left(\mathbb{C}, C_t\right). \tag{9}
$$

Then, we use the following two feature fusion technologies (Sun 2005) to fuse the features:

$$
FFS1: \ Z_1 = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}^T \begin{bmatrix} K\left(m, m_t\right) \\ K_{log}\left(\mathbb{C}, C_t\right) \end{bmatrix} \tag{10}
$$

$$
FFS2: \ Z_2 = \begin{bmatrix} A \\ B \end{bmatrix}^T \begin{bmatrix} K\left(m, m_t\right) \\ K_{log}\left(\mathbb{C}, C_t\right) \end{bmatrix} \tag{11}
$$

Finally, 1-NN classifier in this p-dimensional common subspace is conducted based on Euclidean distance.

Besides, two new feature fusion methods are also proposed, termed as feature fusion strategy 3 (FFS3) and feature fusion strategy 4 (FFS4), respectively.

$$
FFS3: \ Z_1 = \begin{bmatrix} \mu A & 0 \\ 0 & \left(1-\mu\right) B \end{bmatrix}^T \begin{bmatrix} K\left(m, m_t\right) \\ K_{log}\left(\mathbb{C}, C_t\right) \end{bmatrix} \tag{12}
$$

$$FFS4: \ Z_2 = \begin{bmatrix} \mu A \\ \left(1-\mu\right) B \end{bmatrix}^T \begin{bmatrix} K\left(m, m_t\right) \\ K_{log}\left(\mathbb{C}, C_t\right) \end{bmatrix} \tag{13}$$

where $\mu \in \left[0,1\right]$ is the fusion weight. Different from FFS1 and FFS2, the proposed FFS3 and FFS4 methods dynamically fuse two features, which can improve the recognition rate by finding the best fusion ratio. They can be seen as the generalization of FFS1 and FFS2.

4. EXPERIMENTS AND SIMULATIONS

This section includes four parts. The main objective of the first part is to test the effectiveness of our proposed methods in video recognition using the Honda/UCSD database. In the second part, we will use the more challenging database YouTube Celebrities to verify the proposed algorithms. Object recognition on ETH-80 database is performed in the third part. In the last part, we will compare the real time consuming of different video classification methods. All the data we used in this paper are normalized databases. All experiments are carried out on a PC with system configuration Inter(R) Core(TM) 2.60GHz with 8GB RAM.

In what follows, in order to examine the effectiveness of our proposed methods, a series of comparative experiments with discriminant canonical correlations (DCC) (Kim 2007), manifold-manifold distance (MMD) (Wang 2008), covariance discriminative learning (CDL) (Wang 2012), localized multi-kernel metric learning (LMKML) (Lu 2013), Log-Euclidean metric learning (LEML) (Huang 2015) and projection metric learning (PML) (Huang 2015) are performed.

For fair comparison, the important parameters of each method were empirically tuned according to the recommendations in the original references as well as the source codes provided by the original authors. In DCC and MMD, PCA was performed to learn the single or mixture of linear subspaces by preserving 95% of data energy. In CDL, to avoid the matrix singularity, regularization was applied to the original covariance matrix as: $C^* = C + \lambda I$, where I is the identity matrix and λ was set to $10^{-3} \times trace\left(C\right)$, and $c-1$ dimensions discriminant features are used, where c is the number of classes. In LEML, the parameter gamma is searched from the range [0.001,0.01,0.1,1,10,100,1000], zeta is tuned from [0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1]. In PML, the parameter α is set to 0.2.

First, we validate the performance on Honda/UCSD database. The Honda/UCSD database contains 59 video sequences of 20 different subjects. In each video sequence, the head pose and facial expression of subject vary largely, and it contains approximately 300~500 frames. For a fair comparison, we randomly select 20 sequences for training and the remainder 39 sequences for testing. Following the similar settings as in (Wang 2008), the famous Viola and Jones face detector is used to collect face. After this, the detected faces are resized to 20x20 images, and histogram equalization is utilized to reduce the illumination variations. Thus, each video sequence generated a video of faces.

The average classification results and square deviations of ten random experiments are listed in Figure 2. We can see that our approaches perform better than the other compared video classification methods. Especially, all our four proposed methods achieve 100% recognition accuracy.

Second, we evaluate the performance on YouTube Celebrities database. The YouTube Celebrities database is a large scale video database collected by Kim et al. for face tracking and recognition, it contains 1910 video clips of 47 celebrities from YouTube web site. Each clip contains hundreds of frames, and these frames are mostly low resolution and highly compressed, hence, the recognition task becomes much more challenging. The face in each frame is also detected by the Viola-Jones face detector, and resize it to 20×20 scale image. Histogram equalization is also used for each image. In each random experiment, one person had 3 randomly chosen videos for the gallery and 6 for probes. The average classification results and square deviations of ten random experiments are shown in Figure 3.

Figure 2. Average classification results and the square deviations of different methods on the Honda/UCSD database (%).

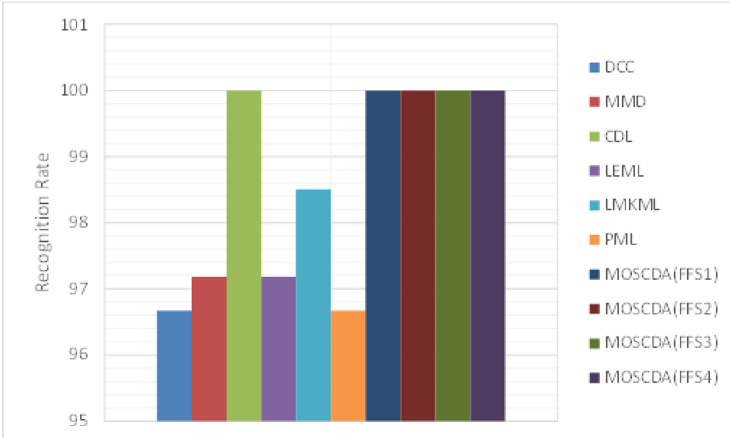
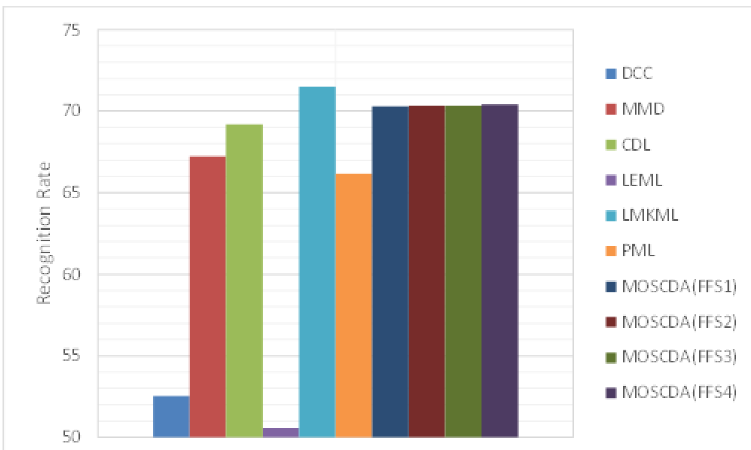


Figure 3. Average classification results and the square deviations of different methods on the YTC database (%).

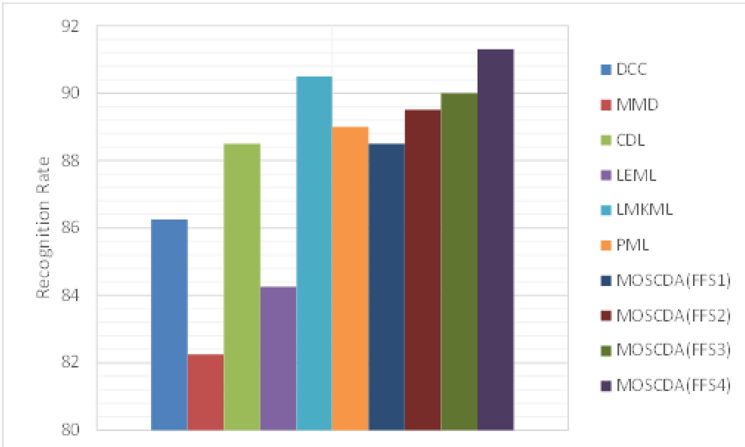


From this figure, we can see that LMKML achieves the highest classification accuracy, while our method MOSCDA(FFS4) achieves the suboptimal recognition rate. The other three MOSCDA methods also achieve better classification accuracy than most state-of-the-art methods. This indicates that our proposed method is effective. Compare our proposed feature fusion methods FFS3, FFS4 with the existing FFS1 and FFS2 methods, our proposed methods perform better, this is because the fact that our method is a generalization of FFS1 and FFS2, which can find the optimal feature fusion proportion.

Third, we evaluate the performance on ETH-80 database. ETH-80 is a classical object category recognition database, it contains visual object images of eight different categories including apples, cars, cows, cups, dogs, horses, pears and tomatoes. In each category, there are 10 object instances (i.e., 10 videos) and each object instance includes 41 different views images. The task is to classify a video of an object instance into a known category. Like previous studies, all object images were segmented from the simple background and then resized them to 20x20 scale for classification. The average classification results and square deviations of ten random experiments are tabulated in Figure 4.

From Figure 4, it can be found that our approach MOSCDA(FFS2), MOSCDA(FFS3) and MOSCDA(FFS4) perform better than the other compared video classification methods, especially

Figure 4. Average classification results and the square deviations of different methods on the ETH-80 database (%).



MOSCDA(FFS4), which achieves the highest recognition rate. Even for our worst method MOSCDA(FFS1), its performance is not worse than CDL, this mean that our information fusion strategy is successful. Compare FFS3 and FFS4 with FFS1 and FFS2 methods, we again find that our proposed FFS3 and FFS4 outperform FFS1 and FFS2.

Forth, we will compare the computational complexity of different video classification methods on Honda/UCSD, YouTube Celebrities and ETH-80. The computing cost (training time + testing time) for each method is tabulated in Table 1.

Table 1. Computation time (seconds) of different methods on different databases.

Methods	Honda/UCSD	YouTube Celebrities	ETH-80
DCC	5.43	59.95	7.99
MMD	69.00	1584	2.09
CDL	12.03	20.03	10.12
LEML	24.17	439.53	45.38
PML	7.09	83.95	12.52
MOSCDA	5.76	18.89	4.23

From this table, we can see that our method MOSCDA costs smaller time than most video classification methods on all databases. Specially, on YouTube Celebrities database, our method consumes the least time; on Honda/UCSD and ETH-80 databases, only DCC or MMD cost less time than our method. At the same time, we can find that the statistics feature based methods (CDL and MOSCDA) are more stable in the aspect of computing time on different databases. Furthermore, compare with CDL, the computing cost of MOSCDA is less than it, even though the mean information of video is added to our method. This is because when we compute the covariance matrices, the mean vectors are also need to be computed, thus, in our method we only need to store them. This also indicate that the computing time of KCCA is less than kernel partial least squares (KPLS) and kernel linear discriminant analysis (KLDA).

5. CONCLUSION

The Web of Things has become a significant symbol of the information age, and the data acquisition is an important basic work in the WoT system. In order to promote the development of data acquisition system, this paper designs a WoT-based enterprise management data acquisition and analysis system. In order to solve the problems of traditional data acquisition systems that cannot respond in real time, occupy storage and data transmission volume is too large, based on canonical correlation analysis (CCA), the multiple order statistics correlation discrimination analysis algorithm is proposed as a fusion strategy. In experiments, we compare our proposed method with DCC, MMD, CDL, LMEL, LMKML and PML. The experimental results show that the proposed MOSCDA approach delivers the overall best performance with the highest recognition rate and comparable computing time in all three databases. However, as new video classification methods, there are still some aspects that deserve study in the future. Firstly, more video representation methods can be considered to jointly model videos. Secondly, we are interested in designing new multi-view learning method. Lastly, the internal variance of a video can also be researched to improve the WoT-based enterprise management data acquisition and analysis system.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Chen, S., & Wang, Y. (2012). Data collection capacity of random-deployed wireless sensor networks under physical models. *Tsinghua Science and Technology*, *17*(5), 487–498. doi:10.1109/TST.2012.6314526

Huang, Z., Wang, R., & Shan, S., Li, X., & Chen, X. (2015). Log-Euclidean metric learning on symmetric positive definite manifold with application to video classification. *Proceedings of the 32nd International Conference on Machine Learning*, 720-729.

Huang, Z., Wang, R., Shan, S., & Chen, X. (2015). Projection metric learning on Grassmann manifold with application to video based face recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 140-149. doi:10.1109/CVPR.2015.7298609

Kim, T., Kittler, J., & Cipolla, R. (2007). Discriminative learning and recognition of video classes using canonical correlations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(6), 1005–1018. doi:10.1109/TPAMI.2007.1037 PMID:17431299

Li, L. (2011). LoT data management and intelligent processing. *ZTE Communications*, *17*(1), 38–41.

Li, Z., Bi, H., & Li, Y. (2009). An adaptive improvement to LRFU replacement strategy. *Computer Engineering and Applications*, *44*(17), 153–157.

Liu, C., Jing, X., & Dong, G. (2011). Talking about the technical characteristics of the Internet of Things and its wide application. *Scientific Consultation, 9*, 86-86.

Liu, Y., & Zhou, G. (2012). Key technologies and applications of internet of things. In *2012 Fifth International Conference on Intelligent Computation Technology and Automation* (pp. 197-200). IEEE. doi:10.1109/ICICTA.2012.56

Lu, J., Wang, G., & Moulin, P. (2013). Video classification using holistic multiple order statistics features and localized multi-kernel metric learning. *IEEE International Conference on Computer Vision*, 329-336.

Sun, Q., Zeng, S., Liu, Y., Heng, P., & Xia, D. (2005). A new method of feature fusion and its application in image recognition. *Pattern Recognition*, *38*(12), 2437–2448. doi:10.1016/j.patcog.2004.12.013

Wang, R., Guo, H., Davis, L. S., & Dai, Q. (2012). Covariance discriminative learning: A natural and efficient approach to video classification. *IEEE Conference on Computer Vision and Pattern Recognition*, 2496-2503.

Wang, R., Shan, S., Chen, X., & Gao, W. (2008). Manifold-manifold distance with application to face recognition based on video. *IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.