International Journal of Smart Home
Vol. 9, No. 9 (2015), pp. 103-112
http://dx.doi.org/10.14257/ijsh.2015.9.9.11

Research on Novel Image Classification Algorithm based on Multi-Feature Extraction and Modified SVM Classifier

Bingquan Huo and Fengling Yin

Binzhou Polytechnic, Shandong, China
huobingquan6@126.com

Abstract

In this research paper, we conduct theoretical analysis and numerical analysis on novel image classification algorithm based on multi-feature extraction and modified SVM classifier. Image object classification and detection are two important basic problems in the study of computer vision, image segmentation, object tracking, behavior analysis and so on the basis of other high-level vision tasks. Existing image classification method can make full use of every single feature between the complementary characteristics of the extracted features of a large number of redundant information, which can lead to image classification accuracy is not high. For this, put forward an improved support vector machine (SVM) based on characteristics and integrated method of image classification. This method can extract comprehensive description of image content features, using principal component analysis to extract the characteristics of transformation, remove redundant information. The experimental result proves the effectiveness and feasibility of the proposed algorithm. In the final part, we conclude the paper and set up the prospect for the future research.

Keywords: *Image Classification, Kernel Function, Modified SVM, Partial Optimization*

1. Introduction

Image object classification and detection are two of the most essential problems in computer vision. They are the basis of many other complex vision problems, such as segmentation, tracking, and action analysis. Object classification and detection is computer vision, pattern recognition and very active research direction in the field of machine learning. Object classification and detection is widely used in many fields, including face recognition, and pedestrian detection in the field of security, intelligent video analysis, the pedestrian tracking, object recognition, vehicle traffic scene in the field of traffic count, retrograde motion detection, license plate detection and recognition, and the Internet in the field of content-based image retrieval, photo album automatic classification and so on. Can say, object classification and detection has been applied to every aspect of People's Daily life, computer automatic classification and detection technology also to a certain extent reduce the burden of people, changed the human way of life. Image classification is a key task in many areas, so how to use the computer quickly and accurately classify the image identifying technology got many scholars pay attention to. In [1], Soheil *et al.* proposed the multimodal task-driven dictionary learning for image classification. Dictionary learning algorithms have been successfully used for both reconstructive and discriminative tasks, where an input signal is represented with a sparse linear combination of dictionary atoms. While these methods are mostly developed for single-modality scenarios, recent studies have demonstrated the advantages of feature-level fusion based on the joint sparse representation of the multimodal inputs. In [2], Rongrong Ji's research group introduced a novel spectral-spatial constraint hyperspectral image classification methodology. In the paper, they propose a hyperspectral image

ISSN: 1975-4094 IJSH

classification method to address both the pixel spectral and spatial constraints, in which the relationship among pixels is formulated in a hypergraph structure. In [3], Anita's group conduct research on feed-forward hierarchical model of the ventral visual stream applied to functional brain image classification. There are more literatures in [4-11] which are focused on the research on image classification with experimental analysis.

Object classification and detection is the basic problem in vision research which is also a very challenging problem. The difficulties and challenges of the object classification and detection in this article could be divided into three levels: the instance level, class level and semantic level. (1) The instance level. On a single object instance, usually due to the light conditions in the process of image acquisition, the different shooting angle, distance, object of non-rigid deformation in itself, and other parts of the barrier, makes the apparent characteristics of the object instance produced very big change, has brought great difficulties to the visual identification algorithm. (2) The class level. Difficulties and challenges often comes from three aspects, the first is a big difference in the class, which is belong to the same class of objects apparent characteristic difference is bigger, the reasons mentioned above all sorts of dynamic change and again is the interference of background, in the actual scenario, the object is not possible in the context of a very clean, often on the contrary, the background can be very complex, the interference existing in the object which we are interested in, making it difficult to identify problems. (3) The semantic level. Difficulties and challenges associated with the visual semantic of image, the level of difficulty is often very difficult to deal with, especially for today's computer vision theory level. A typical problem is called the multiple stability features. Therefore, to solve the mentioned drawbacks and weaknesses, we conduct theoretical and numerical analysis on novel image classification algorithm based on multi-feature extraction and modified SVM classifier in this paper.

2. The Feature Extraction Techniques and Methodologies

Because single feature can only describe part attribute of the image, the image content description is one-sided, lack of enough to distinguish the information, in place of image has great changes usually cannot obtain better classification effect. Therefore, this paper proposes a comprehensive feature extraction method. The underlying characteristics is the first step in object classification and detection framework, there are two kinds of bottom feature extraction methods, one is based on the interest point detection, the other is with the method of concentrated extract. Interest point detection algorithm has a clearly defined through certain standards selection, the obvious features of local texture pixels, edges and corners, blocks, etc., and often able to obtain certain geometrical invariability, thus can get more meaningful in smaller overhead. Use more object classification field in recent years is the concentrated extract, from the image by fixed step length, scale, a large number of local features are extracted, a large number of local description though has higher redundancy, but more abundant information, later after the word package model is used to effectively express usually can get a better performance than interest point detection. Over the years the best object classification algorithm uses a variety of characteristics, sampling methods on extraction combined with interest point detection, the underlying character description also uses a variety of character description, the advantage is, in the feature extraction stage, through to a large number of redundant feature extracting, the maximum to the underlying image description, to prevent the loss of too much useful information, these underlying the redundant information in the description at the back of the main character coding and feature abstraction and degenerate. Widely attention in recent years, in fact, the depth of the learning theory in one of the most important point is that the underlying character description manual design as a first step in visual information processing, tend to be premature loss of useful information, learning directly from the image pixel feature description is than manual

features related to the task more effectively. Single feature vector by merging of comprehensive feature set can be referred to the following equation:

$$\left\{ \left[F_{ip_i}^k, F_{jp_j}^k, ..., F_{sp_s}^k \right]_{1\times\left(p_i+p_j+...+p_s\right)} \middle| 1 \le \mathrm{i} < \mathrm{j} < ... < \mathrm{s} \le \mathrm{m} \right\} \tag{1}$$

Characteristics on the bottom of the concentrated extract contains a large number of redundancy and noise, in order to improve the robustness characteristic expression, you need to use a feature transformation algorithm to encode the underlying characteristics, in order to gain a more robust and distinguish sexual features of expression, this step on the performance of the object recognition has the vital role, and a lot of research work has focused on looking for more powerful feature coding methods. The features selected are shown below.

EEG Features		Num. of Dimension
Zero Crossing Rate		C
Content percentage of the power spectrum	θ wave (4-7Hz)	C
	slow-α wave (7-9Hz)	C
	mid-α wave (9-11Hz)	C
	slow-α wave (11-13Hz)	C
	β wave (13Hz-)	C
Power spectrum of the hemispheric asymmetry [11]	θ wave (4-7Hz)	$2P$
	slow-α wave (7-9Hz)	$2P$
	mid-α wave (9-11Hz)	$2P$
	slow-α wave (11-13Hz)	$2P$
	β wave (13Hz-)	$2P$

Figure 1. The Features Selected for Analysis in the Paper

Vector quantization coding is quantified by a thought, using a smaller feature sets to describe the characteristic of the underlying, achieve the goal of feature compression. Local characteristics in the actual image, the image often has certain fuzziness which represent the local characteristics may and multiple visual word difference is small, this time using vector quantization coding will only use the nearest visual words, while ignoring the other visual word similarity is very high. The following formula shows the sine transform process.

$$F\left(u,v\right)=\frac{2}{N+1}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f\left(x,y\right)\sin\left[\frac{\left(x+1\right)\left(u+1\right)\Pi}{N+1}\right]\sin\left[\frac{\left(y+1\right)\left(v+1\right)\Pi}{N+1}\right] \tag{2}$$

Sparse coding on the object classification's success is not difficult to understand, for a large feature set, an object is usually associated with only few features, for example, bike usually show part such as wheels, handlebar and visual words are closely related, and aircraft, the wing and the TV screen is very small, and pedestrians are usually on the head, limbs and other corresponding visual words have a strong response. Sparse coding one problem, namely, the local characteristics of similar after sparse coding may respond in different visual words, discontinuity of the transformation will generate encoded characteristics do not match, the influence on the performance of the characteristic to distinguish. To solve the issue, we need to combine the PCNN technique to obtain higher accuracy. The PCNN has a structure of 2D array of neurons in which each neuron is connected to a corresponding pixel in the image. In the formula 3~4, we illustrate the steps to solve the connection issue.

$$F_{ij}\left(n\right)=F_{ij}\left(n-1\right)\times\exp\left(-1/AF\right)+A_{ij} \tag{3}$$

$$L_{ij}\left(n\right)=L_{ij}\left(n-1\right)\times\exp\left(-1/AF\right)+VL\times Y_{mn}\left(n-1\right) \tag{4}$$

On a local manifold to encode the underlying characteristics of refactoring, both to ensure to get the characteristics of the discontinuous problems of coding will not have a

sparse coding, and maintain the characteristics of sparse coding sparse. Local linear constraints in the encoding, locality is a core idea of local linear constraint coding, with the introduction of locality, to some extent, improve the continuity of the character encoding process problems, namely the local characteristics of similar distance after coding should still can fall on a local manifold. The figure 2 shows the model flowchart.

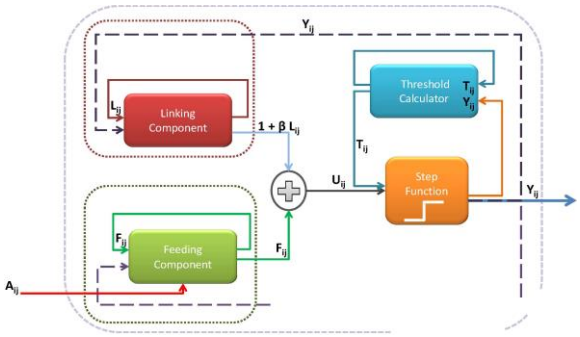

Figure 2. The Basic Model of PCNN

Fisher vector coding at the same time combines the ability of production model and the discriminant model, unlike traditional feature coding method based on reconstruction, which record the local characteristics and visual words first difference and second order difference between. Spatial features together is encoded in the feature set of integrated operation, through the characteristics of the encoded, each d take its maximum or average, get a compact expression feature vector as the image. This step the image expression can get certain feature invariance which also avoids using feature set to the high cost of image expression. Maximum gathering in the vast majority of cases of convergence performance is better than average, is also the most widely used in the classification. No longer use a visual words describe local features, but by the K nearest visual word is described by a weighted, effectively solve the problem of the ambiguity of visual words, improve the precision of object recognition. In recent years is a lot of attention in the field of vision research, the researchers first found that cells in the physiological experiments in most of the time is in the inactive state, namely on the timeline cell activation signal is sparse. Sparse coding by least-square refactoring to join sparse constraint in a complete implementation based on sparse sexual response. For character description is the main research content of object classification. Generally speaking, the object classification algorithm through manual features or learning methods global description of the whole image, and then use the classifier to judge whether there is a certain object. Object detection task is more complex, it needs to be answered in an image where there is the object, so in addition to the feature expression, object structure is the classification of object detection task is different from the place of the most important. The formula 5 shows the feature and the figure 3shos the structure.

$$\left\{ f\left(x_i\right)\right\}_{i=1,2,\ldots,n}, f\left(x_i\right) \in \square^{m} \tag{5}$$

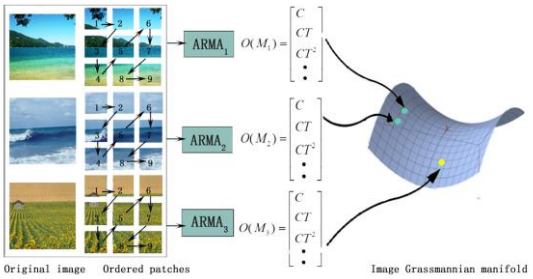

Figure 3. The Structure of the Feature Extraction Process

3. The Proposed Image Classification Methodology

3.1. The Overview of the SVM and Deep Learning Model

We general use support vector machine classifier to classify. After extracted from the image to the characteristics of expression, an image can use a fixed dimension vectors, the next step is to learn a classifier to classify images. This time can choose classifier is a lot of, common classifier with support vector machine (SVM), K-neighbor, neural network and random forest, etc. The support vector machine (SVM) based on maximizing the border is one of the most widely used classifiers, good performance in image classification tasks, especially using the kernel method of support vector machine (SVM). By studying a complete sparse characteristics, can be in high dimensional feature space to improve the characteristics of the linear separation using linear support vector machine (SVM) had the best classification results, greatly reduce the consumption of time and space for the trained classifier. Along with the development of the object classification research, use of visual word size increasing, the image expression dimension is also increasing, reached hundreds of thousands of orders of magnitude. Such a high dimension data, compared with the scale of tens of thousands of data sample, with traditional pattern classification problems have great differences. With continuously increasing scale of processing data, linear classifier based on online learning become the preferred, attracted widespread attention and application.

Automatic encoder is composed of the encoder and decoder, the encoder input data into transformation to express hidden layer and decoder is responsible for from hidden layer to restore the original input. Hidden layer unit number is usually less than the dimension of data input, to play the role of a similar "bottleneck", keep the most important information in data, so as to realize data dimension reduction and feature encoding. First start with raw input for the visible layers, training a single matter, then fixed the first layer of matter weight in matter as a new visual layer, hidden layer unit response training the next layer of matter, and so on. Through this kind of greed unsupervised training, can make whole DBN model to get a better initial values, then add tag information, by means of production or discriminant method, fine adjustment of the whole network supervision, to further improve the network performance. DBN multilayer structure, making it able to learn from hierarchical feature expression, automatic feature abstraction, and unsupervised training process is greatly improved depth of neural network in data quantity is not enough when serious local problems [12-15].

3.2. The Modelling of the Algorithm

The theoretical circle is generally believed that under the condition of the same amount of information, integration of multiple classifiers learning methods usually can get the better generalization ability than a single classifier. Although the algorithm has good generalization ability, but usually contains a large amount of redundant information in image data and the redundant information and greatly damage the generalization ability of learning. In addition, many features combination will result in higher dimension, and the rise of the characteristic dimension of will result in using the algorithm for image classification, the SVM training and testing of occupied time growth, but how many and classification of the characteristic dimension of the effect is not necessarily a link between. Therefore, when using the algorithm for image classification, need to extract the feature dimension reduction, remove the redundant information in the feature. Here we will be the most popular word package model compared with convolution neural network model, found that the two are very similar. In word package models, characteristics of the underlying characteristics of the coding process, in fact the approximate equivalent to the convolution layer in the neural network, and the convergence layer of operation and the convergence of word package model also. The difference is that the word package

effectively only contains a convolution model and a convergence layer, and the model using unsupervised learning characteristics in the form of expression, and the convolutional neural network contains multiple layers of simple and complex cells, can undertake more complex characteristics of the transformation, and the process of the learning process is supervised, filter weights can be adjusted according to the data and tasks, thus learn more meaningful expression characteristics. From this perspective, convolution neural network has the characteristics of the more powerful ability to express its outstanding performance in the image recognition task is easy to explain.

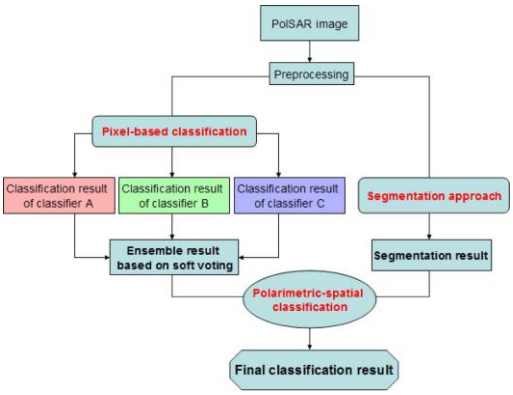

Figure 3. The Flowchart of the Proposed Methodology

The smallest Euclidean distance is shown in the following expression:

$$D_{Grassmann}\left(X_i, X_j\right) = \min_{M>0} \left\| \left(C\left(X_i\right) - C\left(X_j\right)M\right)^T \left(C\left(X_i\right) - C\left(X_j\right)M\right) \right\|_F \tag{6}$$

From the Angle of the model, the object detection mainly adopts is variable component model, more attention to the local characteristics, object classification model is the main model of word in package, from the perspective of the process of both, they used the information is different, object detection more is to use their own information, local information, more is the use of the information of the image object classification, namely global information. They each have advantages and disadvantages, the object of local information to consider more structure information, which makes the object detection and classification accuracy is higher, but also not strong robustness problems of object classification; Global information to consider more global statistical information of an image, especially the semantic information of images, which can consider more information to determine, but the increase of the amount of information may lead to improvements in accuracy, also may be due to redundancy degrade the performance of the classification, but from a statistical sense, its robustness can be improved. Object detection and classification, therefore, between the larger differences, also means that there is a bigger of the two complementary. The following expressions show the model.

$$d\left(\left[T\right]\left|\left[\sum\nolimits_i\right]\right.\right) < d\left(\left[T\right]\left|\left[\sum\nolimits_j\right]\right.\right), \qquad j = 1, 2, ..., M \tag{7}$$

$$D_{Grassmann}\left(X_i, X_j\right) = \left\| I - M^T M \right\|_F \tag{8}$$

Object classification task to determine whether the image contains, global express more key; Object detection task is to determine the location of the objects in the image and scale, object structure is even more crucial. In the era of big data, massive video data from the complicated variability (variability) will bring enormous challenges to the characteristics of the traditional learning method. And deep learning model expression ability, strong natural data will undoubtedly for big data under the background of the visual research produced great impact, is bound to image detection, object classification

research to new heights. There are of course, the current deep learning model explanatory difference, the model complexity is high, the optimization of hard times, the calculation problems of higher intensity, this requires researchers to further thinking. Explicit structure, for example, a priori embedded deep learning model, can effectively reduce the size of the network parameter space and reduce the local minima problem, which can more effectively solve the detection, segmentation and other tasks. After a given model expression, how to learn from a given data model parameters, is a core problem in structured learning. At present there are usually based on the study of probability method, such as maximum likelihood estimation, maximum a posteriori estimation etc. are based on the minimization of a loss function. The efficiency and accuracy of different methods in learning have differences on the fast effective learning algorithm, is of special value. Concentrate hierarchical expression, that is, deep learning. Depth study focused on the expression of study, which is paid more attention to an input to get the corresponding output, the characteristics of the transformation of to lack natural explanation, more like a black box system shown in the figure 4.

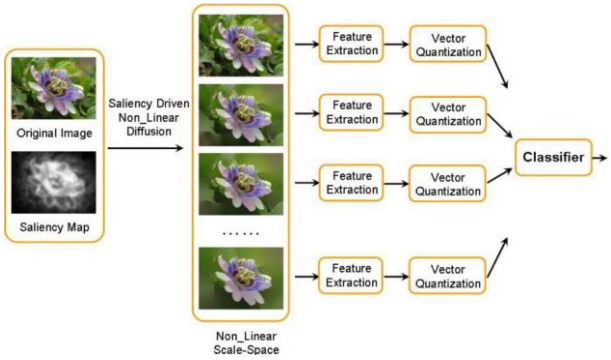

Figure 4. The Proposed Black Box System

4. Experiment and Simulation

4.1. The Experiment Set-up and Initiation

In order to verify the validity of the algorithm in this paper, on the Windows platform, select the Laplacian priori Matlab2011b environment model, TV prior model and GMRF prior model, comparing with the algorithm. The environment is shown below.

Development Language	Java
DBMS	MySQL
Image dataset	Caltech-256 [17]
Number of categories	3
Number of images	96 per category
Number of training images	6 per category

4.2. The Experimental Result

The classification result is shown in the figure 5. In view of the current image classification method has failed to make full use of all kinds of single feature image between complementary characteristics, and the extracted features in the presence of large amounts of redundant information, thus causing the image classification accuracy is not high. We could conclude from the figures that our method performs well.

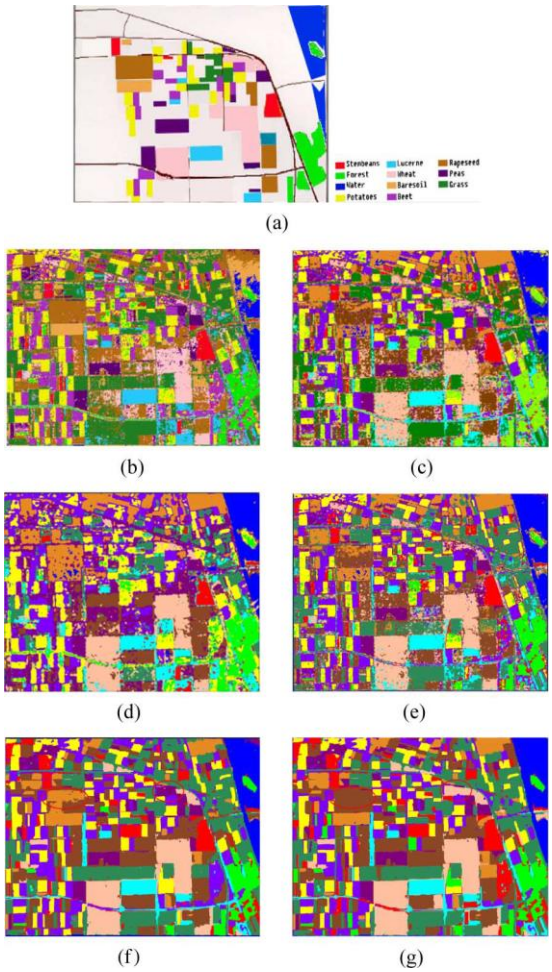

Figure 5. The Experimental Result for the Proposed Algorithm

5. Conclusion and Summary

Image object classification and detection are two of the most essential problems in computer vision. Object classification and detection is computer vision, pattern recognition and very active research direction in the field of machine learning. Object classification and detection is widely used in many fields, including face recognition, and pedestrian detection in the field of security, intelligent video analysis, the pedestrian tracking, object recognition, vehicle traffic scene in the field of traffic count, retrograde motion detection, license plate detection and recognition, and the Internet in the field of content-based image retrieval, photo album automatic classification and so on. Object classification and detection in computer vision research has important theoretical significance and practical application value, at present there are many difficulties and challenges at the same time. On this basis, this paper also discussed the differences and unity of the object classification and detection, object classification and detection of the direction of the further thinking, based on the expression of deep learning from learning and structure proposed in two directions are analyzed. In the future, we plan to conduct more related research to finalize the theoretical analysis and simulation result to obtain better result.

References

[1] S. Bahrampour, N. M. Nasrabadi, A. Ray and W. Kenneth Jenkins, "Multimodal task-driven dictionary learning for image classification", arXiv preprint arXiv, 1502 01094, **(2015)**.
[2] R. Ji, Y. Gao, R. Hong, Q. Liu, D. Tao and X Li, "Spectral-spatial constraint hyperspectral image classification", Geoscience and Remote Sensing, IEEE Transactions on vol. 52, no. 3 **(2014)**, pp. 1811-1824.
[3] D. B. Keator, J. H. Fallon, A. Lakatos, C. C. Fowlkes, S. G. Potkin and A. Ihler, "Feed-forward hierarchical model of the ventral visual stream applied to functional brain image classification", Human brain mapping vol. 35, no. 1, **(2014)**, pp. 38 - 52.
[4] Y. Huang, Z. Wu, L. Wang and T. Tan, "Feature coding in image classification, a comprehensive study", Pattern Analysis and Machine Intelligence, IEEE Transactions on vol. 36, no. 3, **(2014)**, pp. 493-506.
[5] A. Barley and C. Town, "Combinations of Feature Descriptors for Texture Image Classification", Journal of Data Analysis and Information Processing 2014 **(2014)**.
[6] B. Song, J. Li, M. D. Mura, P. Li, A. Plaza, J. M. Bioucas-Dias, J. A. Benediktsson and J. Chanussot, "Remotely sensed image classification using sparse representations of morphological attribute profiles", Geoscience and Remote Sensing, IEEE Transactions on vol. 52, no. 8, **(2014)**, pp. 5122-5136.
[7] H. Wang and J. Wang, "An effective image representation method using kernel classification", In Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on, **(2014)**, pp. 853-858.
[8] J. Li, H. Zhang, Y. Huang and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary", Geoscience and Remote Sensing, IEEE Transactions on vol. 52, no. 6, **(2014)**, pp. 3707-3719.
[9] A. Stumpf, N. Lachiche, J. -P. Malet, N. Kerle and A. Puissant, "Active learning in the spatial domain for remote sensing image classification", Geoscience and Remote Sensing, IEEE Transactions on vol. 52, no. 5, **(2014**, pp. 2492-2507.
[10] L. Guo, and S. Boukir, "Ensemble margin framework for image classification", In Image Processing (ICIP), 2014 IEEE International Conference on, **(2014)**, pp. 4231-4235.
[11] Li, Cheng-Hsuan, H. –H. Ho, B. –C. Kuo, J. –S. Taur, H. –S. Chu and M. –S. Wang, "A Semi-Supervised Feature Extraction based on Supervised and Fuzzy-based Linear Discriminant Analysis for Hyperspectral Image Classification", Appl. Math 9, no. 1L **(2015)**, pp. 81-87.
[12] X. Jiang and D. L. Silver, "A Survey of Transfer Learning in Deep Learning Architectures", In Science Atlantic Mathematics, Statistics and Computer Science Conference **(2014)**.
[13] M. Jiu, C. Wolf, G. Taylor and A. Baskurt, "Human body part estimation from depth images via spatially-constrained deep learning", Pattern Recognition Letters vol. 50, **(2014)**, pp. 122-129.
[14] S. Jia, J. C. Vaughan and X. Zhuang, "Isotropic three-dimensional super-resolution imaging with a self-bending point spread function", Nature photonics, vol. 8, no. 4, **(2014)**, pp. 302-306.
[15] Y. Sun, Y. Chen, X. Wang and X. Tang, "Deep learning face representation by joint identification-verification", In Advances in Neural Information Processing Systems, **(2014)**, pp. 1988-1996.

