A NOVEL REMOTELY SENSED DATA CLASSIFICATION METHOD—MS-SVMS

Dengkui Mo[a*], Hui Lin[a], Jiping Li[b], Zhuo Zang[a], Xiaopo Xue[a],Yujiu Xiong[c]

[a]Research Center of Forestry Remote Sensing and Information Engineering, Central South University of Forestry and Technology, 498 Shaoshan Road, Changsha, Hunan, P. R. China
[b]College of Resource and Environment, Central South University of Forestry and Technology, 498 Shaoshan Road, Changsha, Hunan, P. R. China
[c]College of Resources Science and Technology, Beijing Normal University, Xinjiekou Outer St. 19th, Haidian District, Beijing, R.P. China

KEY WORDS: remote sensing, land cover, classification, segmentation, Landsat.

ABSTRACT:

Support vector machines (SVMs) is a statistical learning method with good performance when the sample size is small, due to their excellent performance, SVMs are now used extensively in pattern classification applications and regression estimation, Unfortunately, it is currently considerably slower in test phase caused by number of the support vectors, which has been a serious limitation for some application such as remotely sensed data classification. To overcome this problem, we introduced mean shift (MS) algorithm to select the feature vectors. Through the MS algorithm, the modes of data are real input vectors and the number of modes is controlled by three physical meaning parameters (h_s , h_r , M). Remotely sensed data has spatial and spectral characters and it has several million pixels in one image generally. Therefore, how to reduce the complexity of the data becomes a crucial problem in remotely sensed data classification based on SVM method. In order to solve such problem, we proposed MS-SVMs classification method. MS-SVMs is the combined process of segmenting an image into regions of pixels based on mean shift algorithm, computing attributes for each region to create objects, and classifying the objects based on attributes, to extract features with SVMs supervised classification. This workflow is designed to be helpful and intuitive, while allowing you to customize it to specific applications. In order to verify the feasibility and effectiveness of proposed method, Landsat ETM image is adopted as original data, and experiments proved the proposed method is robust and efficient, further more, it helps improve classification speed and accuracy observably.

1. INTRODUCTION

With the development of remote sensing technology, the volume of remote sensing imageries increases rapidly, to make use of these huge image data effectively, fast and accurate image analysis algorithms will be very helpful and crucial. In recent years, researchers have developed a lot of image interpretation methods. In general, they can be classified into two categories, one is based on the pixel-based image classification method, and the other is the object-oriented image analysis methods. Traditional pixel-based classification techniques all based on spectral analysis of individual pixels and significant progress has been achieved in recent years, it was limited by utilizing only spectral information. Because of “Same Object With Different Spectra”, “Different Objects With Same Spectrum”, topographical influence, different soil moisture contents, the capability of applying remote sensing data on land cover classification is reduced Each pixel in such classification process have equal status, it did not consider spatial relationships. All these will lead to "Salt and Pepper effect" results. Since the first commercially object-oriented image analysis software——eCognition (Definiens Imaging GmbH) was developed, it has been utilized in many fields successfully (Benfield et al. 2007; Benz et al. 2004; Niemeyer and Canty 2002). Object-oriented image analysis does not operate on single pixels directly; it started with image segmentation in which objects were created using spatial and spectral information. Once objects were defined, the classification was more robust since all pixels of the object were necessarily classified to the same class and the classification result was closer to the human vision.

In this paper, a new object-oriented remotely sensed imagery interpretation method MS-SVMs was proposed. It is mainly composed of robust mean shift—based image segmentation and SVMs classifier. Image segmentation makes classification results closer to the human vision, and after segmentation, the number of objects decreased. It will be significantly reduce heterogeneity in the same land cover type, and increase heterogeneity between different land cover types; it will help improve the SVMs classification speed and accuracy.

2. SUPPORT VECTOR MACHINES CLASSIFICATION

The SVMs is a new and promising classification and regression technique proposed by Vapnik and his group at AT&T Bell Laboratories(Vapnik, 1999). Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. A special property of SVMs is that they simultaneously minimize the empirical classification error and maximize the geometric margin; hence they are also known as maximum margin classifiers.
See Figure 1., viewing the input data as two sets of vectors in an n-dimensional space, an SVM will construct a separating

* Corresponding author, Tel: +86 731 5623137. Email address: dengkuimo@yahoo.com.cn

hyperplane in that space, one which maximizes the "margin" between the two data sets. To calculate the margin, it constructs two parallel hyperplanes, one on each side of the separating one, which are "pushed up against" the two data sets. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the neighbouring data points of both classes. The larger the margin or distance between these parallel hyperplanes, the better the generalization error of the classifier will be.

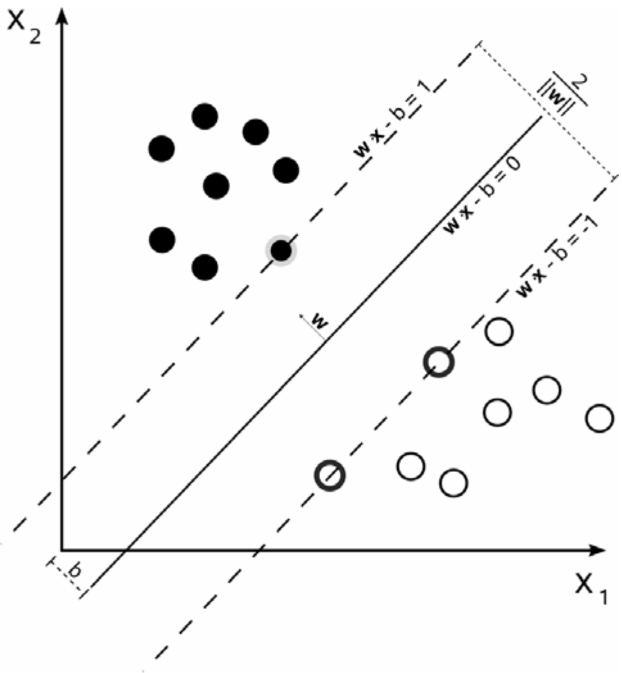

Figure 1. Maximum-margin hyperplane and margins for a SVM trained with samples from two classes. Samples on the margin are called the support vectors.

Given a set of examples $(x_i, y_i), i = 1, \ldots l$ where $x_i \in R^N$ and $y_i \in \{-1, +1\}$, indicating the class to which the point x_i belongs. Each x_i is a p-dimensional real vector. Given the maximum-margin hyperplane which divides the points having $y_i = 1$ from those having $y_i = -1$. Any hyperplane can be written as the set of points x satisfying

$$w \bullet x - b = 0 \tag{1}$$

The vector w is a normal vector: it is perpendicular to the hyperplane. The parameter b determines the offset of the hyperplane from thee origin along the normal vector w. We want to choose the w and b to maximize the margin, or distance between the parallel hyperplanes that are as far apart as possible while still separating the data. These hyperplanes can be described by the equations

$$w \bullet x - b = 1 \text{ and } w \bullet x - b = -1 \tag{2}$$

Note that if the training data are linearly separable, we can select the two hyperplanes of the margin in a way that there are no points between them and then try to maximize their distance. By using geometry, we find the distance between these two hyperplanes is $2/|w|$, so we want to minimize $|w|$. As we also have to prevent data points falling into the margin.

If there exists no hyperplane that can split the "yes" and "no" examples, the Soft Margin method will choose a hyperplane that splits the examples as cleanly as possible, while still maximizing the distance to the nearest cleanly split examples. The method introduces slack variables, ξ_i which measure the degree of misclassification of the datum x_i. The support vector machines require the solution of the following optimization problem:

$$\min_{w,b,\xi} \left(\frac{1}{2} w^T \bullet w + C \sum_{i=1}^{l} \xi_i \right) \tag{3}$$

Subject to $y_i \left(w^T \bullet \varnothing(x_i) + b \right) \geq 1 - \xi_i$, $\xi_i \geq 0$. Here ξ_i are positive slack variables. $C > 0$ is a preset penalty value for misclassification errors. Training vector x_i is mapped into a higher (maybe infinite) dimensional space by the function $\varnothing$. The $w^T \bullet \varnothing(x_i) + b$ is a hyperplane in this higher dimensional space. SVM will find an optimal separating hyperplane. Furthermore, $k(x_i, y_i) = \varnothing(x_i)^T \varnothing(y_i)$ is called the kernel function. The choice of the kernel function k is crucial for good classification performance. There are four types of kernels: linear, polynomial, radial basis function (RBF), and sigmoid. In the study we used the RBF, which works well in most cases. The mathematical representation of RBF kernel:

$$k(x_i, y_i) = \exp\left(-\gamma \bullet \|x_i - y_i\|^2 \right), \gamma > 0 \tag{4}$$

Here, γ is a kernel parameter. The RBF is one of the most commonly used kernel functions. In general, the RBF is a reasonable choice. First, the RBF kernel non-linearly maps samples into a higher dimensional space, so the RBF can manipulate the case when the relationship between class labels and attributes is non-linear. Second, the RBF kernel has less numerical computation difficulties. In recent years, SVMs classification method was applied in remote sensing image interpretation and achieved good results (Dash et al., 2007; Fauvel et al., 2006; Mazzoni et al., 2007). However, there still exists one problem, the huge computation in testing phase caused by the large number of feature vectors, which seriously restricted it into practical applications. We use ENVI 4.2 (www.ittvis.com) software in the experiments. We manually tune the best parameter C and γ on the training set (generally, the system default settings will create a satisfactory result, $C = 0.167$, $\gamma = 100.00$). In practice, we intend to extract six land cover types from remote sensing data (Figure 2.), each category is selected about 1,500 pixels as a training sample, the SVMs training process is very slow, and we will not tolerate such a processing speed. The time taken for SVMs to test a new sample is proportional to the number of the support vectors, so the decision speed will become quite slow if the number of the support vectors is very large. Therefore, how to reduce the complexity of data becomes a crucial problem in SVMs research area (Lardeux et al., 2007). In addition, the boundaries of land cover samples are difficult to confirm, emergence of “Salt and Pepper Effect” and the fault results are difficult to modify. Image segmentation process can divide image into fewer meaningful objects, it will help to solve the above-mentioned problems. We will introduce mean shift-based segmentation algorithm and MS-SVMs method in section 3.

3. MEAN SHIFT-BASED IMAGE SEGMENTATION

3.1 Image segmentation

Image segmentation is one of the low-level computer vision tasks. It refers to the process of partitioning a digital image into multiple regions. The goal of segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyze. The result of image segmentation is a set of regions that collectively cover the entire image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics.

Several general-purpose algorithms and techniques (such as Clustering Methods, Edge Detection Methods, Region Growing Methods, Watershed Transformation etc.) have been developed for image segmentation (Guo et al., 2007). Since there is no general solution to the image segmentation problem, these techniques often have to be combined with domain knowledge in order to effectively solve an image segmentation problem for a problem domain.

3.2 Mean shift algorithm

Image segmentation is misleading difficult. Incorrect results can be easily obtained since the employed techniques often rely upon the user correctly guessing the values for the selecting parameters. To improve performance, the execution of image segmentation tasks should be task driven, i.e., supported by independent high-level information. This approach, however, requires that, first, the low-level stage provides a reliable enough representation of the input and that the feature extraction process is controlled only by very few tuning parameters corresponding to intuitive measures in the input domain. Mean-shift is a non-parametric feature space analysis technique, which can achieve the above-stated goals. Application domains include clustering in computer vision and image processing (such as clustering, tracking, segmentation, discontinuity preserving smoothing, filtering, edge detection, and information fusion etc.) (Islam and Alam, 2007; Li, 2007; Song et al., 2007; Wang and Guo, 2007; Wen and Cai, 2006; Wen and Cai, 2007). The mean shift procedure was originally presented in 1975 by (Fukunaga and Hostetler, 1975). Mean shift is a procedure for locating stationary points of a density function given discrete data sampled from that function. It is useful for detecting the modes of this density. The mean shift feature space analysis has been introduced by (Comaniciu and Meer, 2002).The mean shift procedure is an adaptive local steepest gradient ascent method. The mean shift vector is computed by the following formula:

$$m_{h,G}(\mathrm{X}) = \frac{1}{2} h^2 c \frac{\hat{\nabla} f_{h,K}(\mathrm{X})}{\hat{f}_{h,G}(\mathrm{X})} \tag{5}$$

While $\hat{f}(X)$ is kernel density estimator, h is bandwidth parameter, $K(X)$ and $G(X)$ are kernel functions, The expression (1) shows that, at location X, the mean shift vector $m_{h,G}(\mathrm{X})$ computed with kernel G is proportional to the normalized density gradient estimate obtained with kernel K. The normalization is by the density estimate in X computed with the kernel G. The mean shift vector thus always points toward the direction of maximum increase in the density.

The relation captured in (5) is intuitive; the local mean is shifted toward the region in which the majority of the points reside. Since the mean shift vector is aligned with the local gradient estimate, it can define a path leading to a stationary point of estimated density. The modes of the density are such stationary points. The mean shift procedure obtained by successive

- Computation of the mean shift vector $m_{h,G}(\mathrm{X})$,
- Translation of the kernel $G(X)$ by $m_{h,G}(\mathrm{X})$,

is guaranteed to converge at a nearby point where $m_{h,G}(\mathrm{X}) = 0$.

An image is typically represented as a two-dimensional lattice of $p-$dimensional vectors (pixels). The space of the lattice is known as the spatial domain, while the gray level, color, or spectral information is represented in the range domain. When the location and range vectors are concatenated in the joint spatial-range domain of dimension $d = p + 2$, the multivariate kernel is defined:

$$K_{h_s,h_r}(\mathrm{x}) = \frac{C}{h^2{}_s h^p{}_r} k\left(\left\|\frac{\mathrm{X}^s}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathrm{X}^r}{h_r}\right\|^2\right) \tag{6}$$

where X^s is the spatial part, X^r is the range part of a feature vector, $k(x)$ the common profile used in the two domains, h_s and h_r the employed kernel bandwidths, and C the corresponding normalization constant. The quality of segmentation is controlled by the spatial h_s and the color h_r .

3.3 MS-SVMs method

In order to utilize the spatial and spectral information, reduce the number of samples, make sample selection tasks maneuverable and improve the accuracy of the classification. An object oriented image analysis method – MS-SVMs is proposed.

Let x_i and z_i , $i=1,...,n$,be the d -dimensional input and filtered image pixels in the joint spatial-color domain and L_i the label of the i th pixel in the segmented image.

1. Run the mean shift procedure for the image and store all the information about the d -dimensional convergence point in z_i .
2. Delineate in the joint domain the clusters $\left\{C_p\right\}_{p=1...m}$ by grouping together all z_i which are closer than h_s in spatial domain and h_r in the color domain, i.e., concatenate the basins of attraction of the corresponding convergence points.
3. For each $i=1,...,n$, assign $L_i=\left\{p\left|z_i\in C_p\right.\right\}$.
4. Optional: Eliminate spatial regions containing less than M pixels.
5. Vectorization the segmented image and extraction mean Red, Green and Blue as feature vectors.
6. Sample selection from the vector spatial database as SVMs input feature vectors, training and classification.
7. Manual modification, accuracy assessment, and then creates a new coverage by merging adjacent polygons that have the same land cover type.

4. EXPERIMENTAL STUDY

4.1 Study area and data

The study area lies in Zixing city, which in the Southeastern Hunan Province. It covers approximately 15km × 15km, and the latitude of the study area ranges form 25.991747° N to 26.137354° N and longitude 113.375248° E to 113.520855° E. Landsat ETM image, 512 × 512 pixels was used for the study see Figure 2.

These data are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at the USGS Center for Earth Resource Observation and Science (EROS) in Sioux Falls, South Dakota. http://lpdaac.usgs.gov. The satellite data is cloud free and excellent for extracting land cover information. It was used Band 7 (2.08-2.35 µm), Band 4 (0.76-0.90 µm), Band 2 (0.52-0.60 µm), and the maximum resolution is 30 m per pixel. Six types of land cover (such as water, broadleaf, clear-cut, conifer and town) were extracted by MS-SVMs method and the accuracy was evaluated.

Figure 2. Remote sensing image in study area. False color composition area (Band 7, 4, 2 composition) of the Landsat ETM image of Zixing, Hunan, China, 512×512 pixels in size.

4.2 MS-SVMs image interpretation

The mean shift procedure is not computationally expensive. Careful C++ implementation of the image segmentation algorithm allowed fast processing of the Landsat image. The segmentation with (h_s , h_r , M)=(10,6.5,100) and (h_s , h_r , M)=(12,8.5,250) of the 512×512 color image are shown in Figure 3. Many meaningful objects are obtained. Using a computer with P4 3.0G, RAM 1024 M, Microsoft Windows XP operating system environment, the running time is less than 4 seconds. Practice shows that the segmentation is not very sensitive to the choice of the resolution parameters h_s and h_r , the image segmentation algorithm is robust. Since the control parameter has clear physical meaning, the parameters selection task is controllable. The vector polygons are created and spectral means are extracted and added into the shapefile attribute database during the image segmentation process.

Training samples are selected from vector polygon in shapefile, see Figure 4 (Left). SVMs classifier choose RBF kernel and set parameters ($C=0.167$, $\gamma=100.00$). The classification speed is increased evidently. In such experiment, the whole SVMs classification process consumes 2 seconds including classifier training and classification steps. The image interpretation results see Figure 4 (Right). Use Confusion Matrix to show the accuracy of a classification result by comparing a classification result with ground truth information. (see Table 1.). Overall Accuracy is 97.4689% and Kappa Coefficient is 0.9639. The results mainly rely on the image segmentation precision and the classification accuracy, we can edit polygon edges of inaccuracy or we can emend fault objects in attribute database directly. Finally, we create a new coverage by merging adjacent polygons that have the same land cover type.

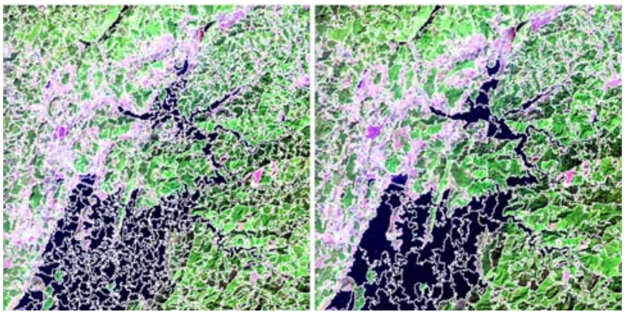

Figure 3. Multiscale image segmentation results based mean shift. Left: 1068 objects, segmentation parameters (h_s , h_r , M)=(10,6.5,100), Right: 391 objects, segmentation parameters (h_s , h_r , M)=(12,8.5,250).

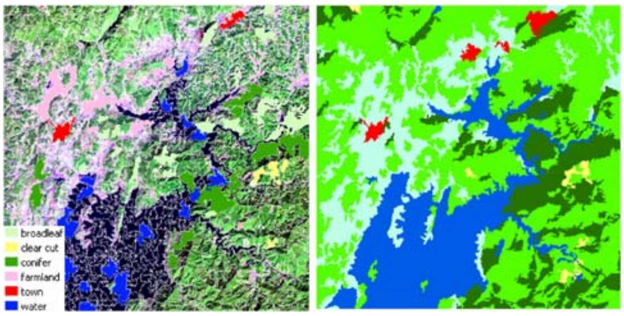

Figure 4. Object oriented sample selection (Left) and classification result (Right).

Class	Ground Truth (Pixel)						
	1	2	3	4	5	6	Total
1	120543	0	0	0	722	0	121265
2	0	50143	0	839	0	0	50982
3	571	0	1684	0	0	0	2255
4	0	0	0	38688	0	0	38688
5	296	0	0	0	45254	0	45550
6	0	0	221	0	446	2737	3404
7	121410	50143	1905	39527	46422	2737	262144

Table 1. Classification accuracy based on proposed method. Replace broadleaf, clear-cut, conifer, farm land, town, water with 1, 2, 3, 4, 5 and 6 respectively.

5. CONCLUSION AND DISCUSSION

Remote sensing image directly reflect the natural surface of the earth, it is complex and ambiguous. How to extract information from remotely sensed imagery quickly and accurately is one of the most important research areas. The MS-SVMs method proposed in this paper is a human-computer interactive object oriented image analysis method and it is not computationally expensive. Since the quality of the output is controlled only by the resolution of the analysis. The results mainly rely on the image segmentation precision which is controlled by three physical meaning parameters (h_s , h_r , M) and the classification accuracy which is directly related to sample selection and classification algorithms design. The MS-SVMs is a general method which is not restricted to the moderate resolution remotely sensed imagery interpretation discussed here. It is suitable for a large variety of remote sensing images interpretation tasks, especially in high spatial resolution remotely sensed imagery.

Otherwise, although this approach is robust and quick in remote sensing interpretation, the mean shift algorithm can also integrate with other supervised classifiers such as maximum likelihood method, minimum distance method and neural network method etc., or the SVMs classifier can also integrate with other robust and fast image segmentation methods. These synthetical algorithms will be evaluated and compared in the future research.

ACKNOWLEDGEMENT

This research was jointly supported by the National Key Technology R&D Program of China (Grant No. 2006BAD23B01), the National Natural Science Foundation of China (Grant No. 30471391), the Natural Science Foundation of Hunan Province, China (Grant No. 02JJBY005) and the Research Foundation of Education Bureau of Hunan Province, China (Grant No. 04B059). These Landsat images are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at the USGS Center for Earth Resource Observation and Science (EROS) in Sioux Falls, South Dakota. http://lpdaac.usgs.gov.

REFERENCES

Comaniciu, D. and Meer, P.,2002. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5): 603-619.

Dash, J. et al.,2007. Land cover classification using multi-temporal MERIS vegetation indices. International Journal of Remote Sensing, 28(6): 1137-1159.

Fauvel, M., Chanussot, J. and Benediktsson, J.A.,2006. Evaluation of kernels for multiclass classification of hyperspectral remote sensing data. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ 08855-1331, United States, Toulouse, France, pp. 813-816.

Fukunaga, K. and Hostetler, L.D.,1975. ESTIMATION OF THE GRADIENT OF A DENSITY FUNCTION, WITH APPLICATIONS IN PATTERN RECOGNITION. IT-21(1): 32-40.

Guo, L., Gao, L., Pian, Z.-Y. and Wang, K.,2007. Image segmentation using multiscale gradient toboggan. ICIEA 2007: 2007 Second IEEE Conference on Industrial Electronics and Applications. Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, Harbin, China, pp. 2206-2209.

Islam, M.M. and Alam, M.S.,2007. Human motion tracking using mean shift clustering and discrete cosine transform. Proceedings of SPIE - The International Society for Optical Engineering. SPIE, Bellingham WA, WA 98227-0010, United States, Orlando, FL, United States, pp. 656616.

Lardeux, C. et al.,2007. Comparison of the ASAR alternative polarisation mode to full polarimetric acquisition for land use estimation over tropical regions. European Space Agency, (Special Publication) ESA SP. European Space Agency, Noordwijk, 2200 AG, Netherlands, Montreux, Switzerland, pp. 6.

Li, P.-H.,2007. Improved mean shift algorithm for object tracking. Zidonghua Xuebao/Acta Automatica Sinica, 33(4): 347-354.

Mazzoni, D., Garay, M.J., Davies, R. and Nelson, D.,2007. An operational MISR pixel classifier using support vector machines. Remote Sensing of Environment, 107(1-2): 149-158.

Song, X., Shen, Z.-K., Wang, P. and Wang, L.-P.,2007. Application of Mean shift method in target tracking. Xi Tong Gong Cheng Yu Dian Zi Ji Shu/Systems Engineering and Electronics, 29(9): 1405-1409.

Vapnik, V.N.,1999. An overview of statistical learning theory. Neural Networks, IEEE Transactions on, 10(5): 988-999.

Wang, K.-J. and Guo, Q.-C.,2007. Image smoothing based on the mean-shift algorithm. Harbin Gongcheng Daxue Xuebao/Journal of Harbin Engineering University, 28(11): 1228-1235.

Wen, Z.-Q. and Cai, Z.-X.,2006. Mean shift algorithm and its application in tracking of objects. Proceedings of the 2006 International Conference on Machine Learning and Cybernetics, ICMLC 2006. Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, Dalian, China, pp. 4024-4028.

Wen, Z.-Q. and Cai, Z.-X.,2007. Convergence analysis of mean shift algorithm. Ruan Jian Xue Bao/Journal of Software, 18(2): 205-212.