ON THE CONVERGENCE AND APPLICATIONS OF MEAN SHIFT TYPE ALGORITHMS

Youness Aliyari Ghassabeh, Tamás Linder, Glen Takahara

Mathematics & Engineering Program, Dept. of Mathematics and Statistics,
Queen's University, Kingston, Ontario

ABSTRACT

Mean shift (MS) and subspace constrained mean shift (SCMS) algorithms are iterative methods to find an underlying manifold associated with an intrinsically low dimensional data set embedded in a high dimensional space. Although the MS and SCMS algorithms have been used in many applications related to information and signal processing, a rigorous study of their convergence properties is still missing. This paper aims to fill some of the gaps between theory and practice. We present theoretical results about convergence of the MS and SCMS algorithms. As well, we discuss potential applications of the SCMS algorithm as a preprocessing step for noisy source vector quantization and nonlinear dimensionality reduction with noisy observations.

***Index Terms*—** Mean shift algorithm, subspace constrained mean shift algorithm, noisy source vector quantization, nonlinear dimensionality reduction

1. INTRODUCTION

Dimensionality reduction and manifold learning are two important problems in many fields related to information processing, including data compression, signal processing, information retrieval, statistical pattern recognition, and artificial intelligence. Real world data, such as speech signals, digital images, genomic data, and fMRI scans, often have high dimensionality that makes their processing difficult and time consuming. It is often desirable that the observed high dimensional data be represented in a lower dimensional space while preserving the original information as much as possible. Dimensionality reduction techniques provide compact and meaningful representations which facilitate compression, classification, and visualization of the high dimensional data. One approach to achieve these goals is to assume that the data of interest lie on or near a low dimensional manifold, embedded in a high dimensional space. This assumption is realistic in many applications where the observed high dimensional data have an intrinsically low dimensional structure. The problem becomes more complicated when the high dimensional input data is corrupted by noise. In this case applying common linear/nonlinear dimensionality reduction techniques [1] on the noisy observations may not lead to a meaningful low dimensional representation of the observed data. To overcome this problem, principal curves and surfaces [2] [3] [4] have been proposed to estimate the low dimensional embedded manifold. A promising class of algorithms, which we refer to as mean shift type algorithms, move the data points in an iterative manner to a low dimensional manifold that can be thought of as principal curves or surfaces. The original mean shift (MS) algorithm [5] iteratively computes modes of an estimated probability density function (pdf). The modes of a pdf play an important role in many machine learning applications, such as classification, image segmentation [5], and object tracking [6]. The collection of these modes can be viewed as a zero-dimensional principal manifold. The recently introduced subspace constrained mean shift (SCMS) algorithm [4] is used to estimate principal curves and surfaces with potential applications in time-varying MIMO channel equalization and time series signal denoising.

Although the MS and SCMS algorithms seem to work well in some applications, a rigorous theoretical study of their convergence properties seems to be missing in the literature. This paper aims to fill some of the gaps between theory and practice, in particular by studying some convergence properties of these algorithms. We show the convergence of the MS algorithm when the stationary points of the underlying kernel density estimate are isolated and propose a sufficient condition for isolated stationary points. We also prove the distance between two consecutive members of the sequence generated by the SCMS algorithm converges to zero which can be used as a stopping criteria for the SCMS algorithm. Furthermore, we consider two applications of the SCMS algorithm: as a preprocessing step for noisy source vector quantization and nonlinear dimensionality reduction for noisy observations.

2. MEAN SHIFT AND SUBSPACE CONSTRAINED MEAN SHIFT ALGORITHMS

2.1. Mean shift (MS) algorithm

The MS algorithm is a non-parametric, iterative method introduced by Fukunaga and Hostetler [7] for locating modes of an estimated probability density function (pdf). It was generalized by Cheng [8] and became popular in the computer vision community where its potential uses for feature space analysis were studied [5]. The MS algorithm iteratively shifts each

This research was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

data point to a weighted average of neighboring points to find stationary points of the estimated pdf. It has been successfully used in many applications ranging from image segmentation [9] to object tracking [10] and information fusion [11]. Let $\mathbf{x}_i \in \mathbb{R}^D, i = 1, \ldots, n$ be a sequence of n independent and identically distributed (iid) random variables. Assume radially symmetric kernel K defined by $K(\mathbf{x}) = c_{k,D}k(\|\mathbf{x}\|^2)$, where $c_{k,D}$ is a normalization factor and $k : [0, \infty) \to [0, \infty)$ is the differentiable profile of the kernel. The estimated pdf using the profile k and bandwidth h is given by [12]

$$\hat{f}_{h,k}(\mathbf{x}) = \frac{c_{k,D}}{nh^D} \sum_{i=1}^{n} k\big(\|\frac{\mathbf{x} - \mathbf{x}_i}{h}\|^2\big). \tag{1}$$

Taking the gradient of (1) and equating it to zero reveals that modes of the estimated pdf are fixed points of the function

$$\mathbf{m}_{h,g}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\big(\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\|^2\big)}{\sum_{i=1}^{n} g\big(\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\|^2\big)} - \mathbf{x}, \tag{2}$$

where $g(x) = -k'(x)$. The vector $\mathbf{m}_{h,g}(\mathbf{x})$ is called mean shift vector [5]. The MS algorithm initializes MS vectors to be one of the observed data. The mode estimate $\mathbf{y}$ in the jth iteration is updated as

$$\mathbf{y}_{j+1} = \mathbf{y}_j + \mathbf{m}(\mathbf{y}_j). \tag{3}$$

The MS algorithm iterates this step until the norm of the difference between two consecutive mode estimates becomes less than some predefined threshold. Although the MS algorithm has been used in different applications, a rigorous proof for the convergence of the algorithm has not been given. For example, a crucial step in the proof given in [5] for the convergence of mode estimate sequence is not correct. In another work [13], it was shown that the MS algorithm is an expectation maximization (EM) algorithm and hence the sequence $\{\mathbf{y}_j\}$ converges to the modes of the estimated pdf. However, the EM algorithm may not converge (e.g., a counterexample is given in [14]), so the convergence of the MS algorithm does not follow.

2.2. Subspace constrained mean shift (SCMS) algorithm

A new definition of a d-dimensional principal surface in $\mathbb{R}^D$ is given by Ozertem and Erogmus [4] as the set of points that are local maximum of a pdf in a local orthogonal $D - d$-dimensional subspace. They proposed the SCMS algorithm [4] to find points that satisfy that definition. Similar to the MS algorithm, the SCMS algorithm starts from data points. It evaluates the MS vector for every data point and computes the mode estimate similarly to the MS algorithm. To estimate points on the d-dimensional principal surface, it projects the MS vectors to the subspace spanned by the $D - d$ eigenvectors corresponding to the $D - d$ largest eigenvalues of the local covariance matrix. The local covariance matrix is defined by [4] $\mathbf{\Sigma}^{-1}(\mathbf{x}) = -\mathbf{H}(\mathbf{x})\hat{f}(\mathbf{x})^{-1} + \mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^t\hat{f}(\mathbf{x})^{-2}$, where $\hat{f}(\mathbf{x})$ is the pdf estimate at $\mathbf{x}$, and $\mathbf{H}(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$ are the Hessian and gradient of the pdf estimate at $\mathbf{x}$, respectively.

The above procedure is iterated until the norm of the difference between two consecutive projections becomes less than a predefined threshold. In [4] simulations for the SCMS algorithm were presented which demonstrated the ability of this algorithm in finding principal curves and surfaces. As well, applications of the SCMS algorithm to time-varying MIMO channel equalization and time series signal denoising were discussed. However, the convergence of the algorithm has not been proved. In the next section, we show that the distance between two consecutive members of the sequence generated by the SCMS algorithm converges to zero. We also propose novel applications of the SCMS algorithm as a preprocessing step for the noisy source vector quantization and nonlinear dimensionality reduction with noisy observations.

3. CONVERGENCE RESULTS

It is proved in [5] that if the kernel K has a convex, differentiable, and strictly decreasing profile k, then the sequence $\{\hat{f}_{h,k}(\mathbf{y}_j)\}_{j=1,2,\ldots}$ is monotonically increasing and thus converges. The authors of [5] claimed that the mode estimate sequence $\{\mathbf{y}_j\}_{j=1,2,\ldots}$ is a Cauchy sequence, which is not true in general. This result was also observed by [15]. For the sequence $\{\mathbf{y}_j\}_{j=1,2,\ldots}$ defined in (4) the following inequality is proved in [5]:

$$\hat{f}_{h,k}(\mathbf{y}_{j+1}) - \hat{f}_{h,k}(\mathbf{y}_j) \geq \frac{c_{k,D}}{nh^{D+2}} \|\mathbf{y}_{j+1} - \mathbf{y}_j\|^2 \sum_{i=1}^{n} g(\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\|^2).$$

If $k(x)$ is a convex and strictly decreasing function such that $0 < |k'(x)|$ for all $x \geq 0$, then $g(x) = -k'(x)$ is always positive. Let $M(j) = \min\{g\Big(\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\|^2\Big), i = 1, \ldots, n\}$. Since $\mathbf{y}_j$ lies in the convex hull $\mathcal{C}$ of the data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, we have $M(j) \geq g(\frac{a^2}{h})$, where a is the diameter of $\mathcal{C}$. Let $\varphi = g(\frac{a^2}{h})$. Hence, the above equality implies

$$\hat{f}_{h,k}(\mathbf{y}_{j+1}) - \hat{f}_{h,k}(\mathbf{y}_j) \geq \frac{c_{k,D}}{h^{D+2}} \|\mathbf{y}_{j+1} - \mathbf{y}_j\|^2 \varphi. \tag{4}$$

Therefore, we have

$$\big(\hat{f}_{h,k}(\mathbf{y}_{j+1}) - \hat{f}_{h,k}(\mathbf{y}_j)\big) \frac{h^{D+2}}{\varphi c_{k,D}} \geq \|\mathbf{y}_{j+1} - \mathbf{y}_j\|^2 \geq 0. \tag{5}$$

Since $\hat{f}_{h,k}(\mathbf{y}_{j+1})$ is a convergent sequence, the limit of the left side of the above inequality converges to zero. Therefore, the following limit relation holds

$$\lim_{j \to \infty} \|\mathbf{y}_{j+1} - \mathbf{y}_j\| = 0. \tag{6}$$

According to the definition of the mean shift vectors, the sequence $\{\mathbf{y}_j\}_{j=1,2,\ldots}$ is in the convex hull of the data set, i.e. $\mathbf{y}_j \in \mathcal{C}, j = 1, 2, \ldots$. Therefore $\{\mathbf{y}_j\}$ is a bounded sequence, satisfying the above limit. But the last two properties are not enough to guarantee the convergence of $\{\mathbf{y}_j\}$. Assuming that the stationary points are isolated, then the total number of the stationary points inside the convex hull of the data points is finite. Note that all stationary points of the estimated pdf are inside the convex hull of the data set and the gradient of the estimated pdf is nonzero outside the convex hull. Now, we have the following theorem for the convergence of the MS sequence.

Theorem 1. *Let $\mathbf{x}_i \in \mathbb{R}^D, i = 1, \ldots, n$. Assume the stationary points of the estimated pdf are isolated. Then the mode estimate sequence $\{\mathbf{y}_j\}_{j=1,2,\ldots}$ converges.*

Proof. The proof of the theorem is quite straightforward in view of (4) and (6) and is omitted; see also [15]. □

The following corollary for the one dimensional case follows from Theorem 1.

Corollary 1. *Let $x_i \in \mathbb{R}, i = 1, \ldots, n$, then the mode estimate sequence $\{y_j\}_{j=1,2,\ldots}$ generated using the Gaussian kernel $k(x) = \exp(-x^2)$ converges.*

Proof. The derivative of the estimated pdf $\hat{f}'_{h,k}(x)$ is a real analytic function, defined on $\mathbb{R}$. Then, either $\hat{f}'_{h,k}(x)$ is a constant function, or the set $S = \{x : \hat{f}'_{h,k}(x) = 0\}$ has not limit points. However, S is clearly a bounded set and $\hat{f}'_{h,k}(x)$ is not a constant function, so S is finite. Then Theorem 1 implies the statement. □

A well-known theorem of differential geometry [16] states that if the Hessian matrix at the stationary points is of full rank, the stationary points are isolated. The following lemma provides sufficient condition for the Hessian matrix to be full rank at the stationary points.

Lemma 1. *Let $\mathbf{x}_i \in \mathbb{R}^D, i = 1, \ldots, n$. Let $\|\mathbf{x}_{max}\|^2$ denote the largest norm among all $\mathbf{x}_i, i = 1, \ldots, n$. Let $\hat{f}(\mathbf{x})$ denote the estimated pdf using the Gaussian kernel with the covariance matrix $\mathbf{\Sigma}$, and let λ_{min} denote the smallest eigenvalue of $\mathbf{\Sigma}$. If $\|\mathbf{x}_{max}\|\lambda_{min} < 1$, then the Hessian matrix of the estimated pdf at the stationary points is of full rank and the stationary points are isolated.*

The following theorem asserts the similar results for the SCMS algorithm.

Theorem 2. *Let $\mathbf{x}_i \in \mathbb{R}^D, i = 1, \ldots, n$. Let $\{\mathbf{y}_j\}_{j=1,2,\ldots}$ be the sequence generated by the SCMS algorithm. If kernel K has a convex and monotonically decreasing profile k, then the sequence $\{\hat{f}_{h,k}(\mathbf{y}_j)\}_{j=1,2,\ldots}$ is monotonically increasing and convergent and $\lim_{j\to\infty} \|\mathbf{y}_{j+1} - \mathbf{y}_j\| = 0$.*

Theorem 2 states that the norm of the distance between consecutive members of the sequence generated by the SCMS algorithm gradually decreases and thus it theoretically satisfies the usual stopping criterion for the SCMS algorithm. The proofs of the preceding two results are omitted due to space constraints.

4. APPLICATIONS

4.1. Noisy source vector quantization

Vector quantization is an important building block used in lossy data compression. It encodes (maps) points of a multidimensional vector space into a finite set of points from the same space called the codebook. The design of quantizers has been widely studied over the past decades and it has been shown that an optimal quantizer of a given codebook size has to satisfy the Lloyd-Max conditions [17]. The Linde-Buzo-Gray (LBG) algorithm [17] satisfies necessary conditions for the optimality of a quantizer and can be used to derive a near optimal codebook. In practice, the source output may be corrupted by noise due to e.g., measurement errors [19]. The structure of the optimal noisy source quantizer for mean square distortion was studied, e.g., by Wolf and Ziv [18]. They showed that to minimize the mean square distortion with respect to the clean data, one needs to quantize the conditional expectation of the clean data given the noisy data [18]. Thus, in order to minimize the distortion, we need to find a good approximation (in the MMSE sense) of the clean data from the observed noisy data. In practical situations where the statistics of the data and noise are unknown, the clean data can be estimated by applying techniques, such as kernel regression, based on training data. However, in practice the availability of training data is not always a realistic assumption and in many situations the designer of the quantizer only has access to the noisy observations, not to the clean source data or any side information. We propose to use the output of the SCMS algorithm as an estimate for the conditional expectation of the clean data given the noisy data. When the data set of interest has an intrinsically low dimensional structure but are corrupted by some noise, the SCMS algorithm can be used to relocate the noisy points onto the low dimensional manifold that supports the clean data. In other words, the SCMS algorithm is used as a preprocessing step and the output of the SCMS algorithm is given to the LBG vector quantizer. Note that the SCMS algorithm only uses the noisy data. It is of interest to numerically evaluate how well the SCMS algorithm approximates the clean data and compare its performance with that of the kernel regression method (which is near optimal if the training data set is large enough). To this end, we compare the mean square distortion that results from quantizing the output of the SCMS algorithm with the near optimal distortion resulting from quantizing the estimated clean data using the kernel regression method. In our

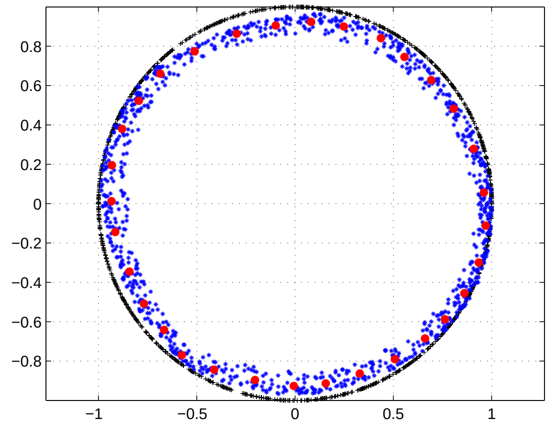

Fig. 1: The black points represent the clean data, the blue points are the output of the SCMS algorithm. The output of the SCMS algorithm is given to the LBG quantizer and the red points are codewords computed using the LBG vector quantization algorithm.

Table 1: The mean square distortion resulted from quantization of the output of the SCMS algorithm and near optimal mean square distortion for different number of codewords ranging from 2 to 128.

Number of the codevectors	2	4	8	16	32	64	128
Optimal distortion [1]	0.7271	0.3858	0.2016	0.1064	0.0595	0.0367	0.0294
SCMS algorithm	0.7274	0.3945	0.2120	0.1220	0.071	0.0498	0.0379

simulations, 1024 points are selected uniformly on the unit circle and then perturbed by bivariate Gaussian noise with independent, zero mean components. The SCMS algorithm accepts the noisy data and produces an estimation of the clean data. The output of the SCMS algorithm is quantized with a quantizer that is trained using the LBG algorithm. Fig. 1 shows the codewords resulting from the LBG algorithm for 32 codewords. The black points in Fig. 1 represent the clean data, the blue points represent the output of the SCMS algorithm, and the red points represent the codewords computed by the LBG algorithm. We vary the number of codewords and run the algorithm for 1, 2, 4, 8, 16, and 32 codewords. Fig. 2 shows the output of the SCMS algorithm and the computed codewords for each simulation. Table 1 compares the mean square distortions resulting from the quantization of the estimated clean data using the SCMS algorithm and the kernel regression method, respectively, as a function of the number of the codevectors (ranging from 2 to 128). Although we do not have access to the training data when using the SCMS algorithm, the simulation results indicate that the mean square distortion resulting from quantization of the output of the SCMS algorithm is close to the near optimal mean square distortion obtained by quantizing the estimated clean data using the kernel regression method.

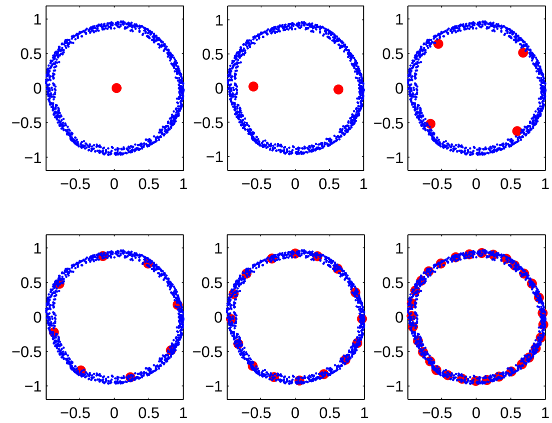

Fig. 2: Applying the SCMS algorithm on noisy data in order to estimate the clean data for noisy source vector quantization. The simulation is done for 1, 2, 4, 8, 16, and 32 codewords. The blue points represent the output of the SCMS algorithm and the red points are the codewords generated by the LBG vector quantization algorithm.

4.2. Nonlinear dimensionality reduction for noisy observations

Often it is reasonable to assume that the observed data set has an intrinsically low dimensional structure but are corrupted by some noise. In this case applying common nonlinear dimensionality reduction techniques such as locally-linear embedding (LLE) [20], kernel principal component analysis (kernel PCA) [21], or Isomap [22] algorithms on the noisy observations may not lead to a meaningful low dimensional representation of the data. When the observed noisy data are not located on an underlying manifold of interest, before applying a dimensionality reduction technique we need first to estimate the points on the underlying manifold. Principal curves and surfaces provide a reasonable low dimensional representation of data and can be used as a preprocessing step before applying common nonlinear dimensionality reduction techniques. As mentioned before, the SCMS algorithm has capability to estimate principal curves and surfaces. After this estimation step, one can apply dimensionality reduction techniques to obtain a representation of the data in a low dimensional space. To illustrate how this works we selected 500 samples uniformly from a three dimensional spiral. Then independent, three dimensional zero mean Gaussian noise with per component variance 0.7 is added to those samples. Fig. 3 shows a scatterplot of the noisy observations and the output of the SCMS algorithm. To assess the performance of the SCMS algorithm, we selected 12 clean data points from the spiral (the markers in Fig. 3 represent the selected points) and applied the LLE algorithm [20] to them to obtain a one-dimensional representation.The second row in Fig. 4 shows the representation of the selected clean points in one dimensional space. Then we applied the LLE algorithm to the estimates of these points computed as the output of the SCMS algorithm that was run on the noisy data. The first row in Fig. 4 shows the representation of the estimated points (output of the SCMS algorithm) after applying the LLE algorithm. The third row in Fig. 4 is one dimensional representation of the noisy points after directly applying the LLE algorithm. It can be observed from Fig. 3 and Fig. 4 that although the observed data was corrupted by Gaussian noise, applying the LLE algorithm to the output of the SCMS algorithm gives a one-dimensional representation very similar to that of the clean data. On the other hand, applying the LLE algorithm directly to the noisy

[1] It is near optimal distortion since the kernel estimate converges to the desired conditional expectation in the limit of large training set sizes. Also the LBG algorithm converges to a local minimum, but is not guaranteed to reach the global minimum.

version of the observed data changes pairwise distances and the one-dimensional order of the points that is not desirable.

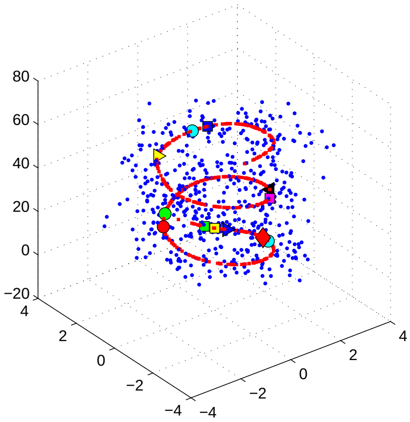

Fig. 3: Applying the SCMS algorithm on the noisy data in order to estimate the clean data.The red points represent the output of the SCMS algorithm and the blue points are the noisy data.

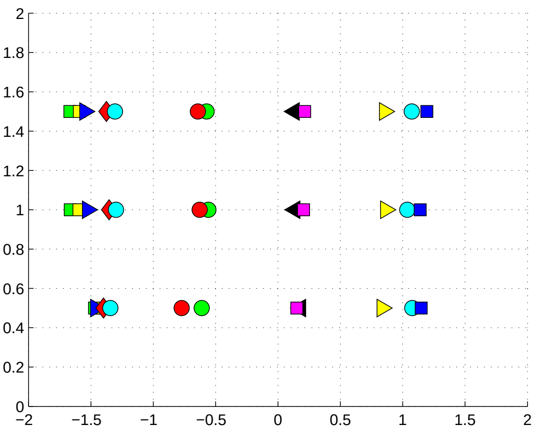

Fig. 4: The first row shows the output of the LLE algorithm when applied to the SCMS estimates of the selected points.The second row shows the output of the LLE algorithm applied to the clean data points. The third row shows the output of the LLE algorithm applied directly to the noisy data points.

5. REFERENCES

[1] J. A. Lee, M. Verleysen, *Nonlinear Dimensionality Reduction,* Springer-Verlag, 2007.

[2] T. Hastie, W. Stuetzle, "Principal curves," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502-516, 1989.

[3] B. Kegl, A. Krzyzak, T. Linder, K. Zeger, "Learning and design of principal curves," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 281-197, Mar. 2000.

[4] U. Ozertem, D. Erdogmus, "Locally defined principal curves and surfaces," *Journal of Machine Learning Research*, vol. 12, pp. 1249-1286, Apr. 2011.

[5] D. Comanicio, P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 603-619, May 2002.

[6] A. Yilmaz, "Object tracking by asymmetric kernel mean shift with automated scale and orientation selection," *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Minnesota, USA, pp. 18-23, Jun. 2007.

[7] K. Fukunaga, L. D. Hostetler, "Estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. on Inform. Theory*, vol. 21, pp. 32-40, Jan. 1975.

[8] Y.Cheng, "Mean shift, mode seeking and clustering," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 790-799, Aug. 1995.

[9] J.Wang, B. Thiesson, Y. Xu, M. Cohen, "Image and video segmentation by anisotropic kernel mean shift," *In Proc. European Conf. on Computer Vision*, Prague, Czech Republic, vol. 2, pp. 238-250, 2004.

[10] C. Jian, Y. Jie, "Realtime Infra-red Object Tracking Based on Mean Shift," *Lecture Notes in Computer Science, Springer,* vol. 3287, pp. 484-495, 2004.

[11] H. Chen and P.Meer, "Robust fusion of uncertain information," *IEEE Trans. Systems, Man, ans Cybernetics-Part B: Cybernetics,* vol. 35, pp. 578-586, May 2005.

[12] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* , Chapman and Hall, 1986.

[13] M. A. Carriera Perpinan, "Gaussian mean shift is an EM algorithm," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 767-776, May 2007.

[14] R. A. Boyles, "On the convergence of the EM algorithm " *Journal of the Royal Statistical Society: Series B*, vol. 45, pp. 47-50, Jan. 1983.

[15] X. Li, Z. Hu, F. Wu, "A note on the convergence of the mean shift," *Pattern Recognition*, vol. 40, pp. 1756-1762, Jun. 2007.

[16] M. Golubitsky, V. Guilemin, *Stable Mappings and their Singularities*, Springer-Verlag, 1973.

[17] Y. Linde, A. Buzo, R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. on Communications*, vol. 28, pp. 702-710, Jan. 1980.

[18] J. K. Wolf, J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. on Information Theory*, vol. 16, pp. 406-411, Jul. 1970.

[19] D. Rebollo-Monedero, S.Rane, B. Girod, "Wyner-Ziv quantization and transform coding of noisy sources at high rates," *Eighth Asilomar Conf. on Signals, Systems, and Computers*, pp 2084 -2088, Mar. 2005.

[20] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323-2326, Dec. 2000.

[21] B. Schölkopf, A. Smola, K.-R. Muller, *Advances in Kernel Methods-Support Vector Learning*, MIT Press Cambridge, 1999.

[22] J. B. Tenenbaum, V. de Silva, J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319-2323, Dec. 2000.