Proceedings of the World Congress on Engineering and Computer Science 2014 Vol II
WCECS 2014, 22-24 October, 2014, San Francisco, USA

The Efficiency of Product Multivariate Kernels

Francis O. Oyegue* and Sunday M. Ogbonmwan

***Abstract*-The Mean Integrated Squared Error (MISE) is a measure of discrepancy between the estimated and true density in kernel density estimation. A more global measure is the Asymptotic Mean Integrated Squared Error (AMISE). This measure (AMISE) is used to quantify the performance of the estimator. However, the focus of this paper is to obtain the efficiency values of some symmetric beta kernels. This is necessary in the sense that it enables one to choose an appropriate kernel, especially in the multivariate setting. We derive formulas to generalize the AMISE and the efficiency. The efficiencies are obtained by taking the ratio of the product (multivariate) kernels considered and the Epanechnikov kernel. This kernel (Epanechnikov kernel) form the basis for the optimum kernel. The results reveal reduction in efficiencies of the beta kernels as their dimension increases.**

***Index Terms*-Density estimation, product kernel, efficiency, Asymptotic mean integrated square error.**

I. INTRODUCTION

Density estimation is simply the construction of an estimate $\hat{f}$ of an underlying density function f for a random variable X drawn from an observed data set. To estimate unknown density estimation, we use either the parametric or the nonparametric methods. The parametric methods such as the maximum likelihood method require the imposition of a functional form on an unknown density. This leads to the problem of the estimation of the parameters.

Sometimes, when the density estimation is unknown and no additional information about the distribution is given, then the nonparametric density estimation, like the histogram or the kernel estimator is applied. This approach allows the data to speak for itself. Instead of the imposition of restrictive parametric assumptions about the underlying distribution, the nonparametric methods allow one to directly approximate the *d*- dimensional density that describes how variables interact [13]. The nonparametric methods are flexible and computationally intensive. The trauma associated with the tedious computations in the nonparametric approach has been considerably reduced via the advent of easily fast computing power in the twentieth century [6]. In this work, we concentrate on one class of nonparametric density estimators, namely, the kernel density estimator. The kernel density estimator is a more reliable statistical technique that deals with some of the problems associated with histogram which are discussed in [2], [10], [20]. In recent time, kernel density estimation has found relevance in huge computational requirement for large-scale analysis [15], [23], and in the area of human motion tracking or pattern recognition [3], [11], [18].

A common term in kernel density estimation is the bandwidth or window width which is analogous to the bin width in histogram. The bandwidth determines how much smoothing is done. Generally, a narrow bandwidth implies that more points are allowed and this lead to a better density estimate. This technique, sometimes , called the Parzen density estimation, was studied in the seminal paper [16], [17], although , the basic idea was independently discussed in [1], [7].

For a d - variate random variable $X_1, X_2, \ldots, X_d$ drawn from a density f the generalized kernel estimation is given as [5]:

$$\hat{f}(\mathbf{x};H)=\frac{1}{n\det H}\sum_{i=1}^{n}K\left(H^{-1}(\mathbf{x}-\mathbf{X}_i)\right) \qquad (1.1)$$

where, $\mathbf{x}=(x_1,x_2,\ldots x_d)^T$ and $\mathbf{X}_i=(x_{i1},x_{i2},\ldots x_{id})^T, i=1,2,\ldots,n$ In this case $K(\bullet)$ is assumed to be the multivariate (*d* – dimensional) kernel. This kernel is assumed to be a product (multiplicative) symmetric probability density function. The scope of the paper is limited to the multivariate kernels that are independent, and supported on a rectangular region. H is the bandwidth matrix which is symmetric and positive – definite. The scaled and unscaled kernels are related by $K_H(\mathbf{x})=|H|^{-\frac{1}{2}}K\left(H^{-\frac{1}{2}}\mathbf{x}\right)$ [21].

An equal bandwidth h in all directions as in (1.1) corresponds to $H=h^2 I_d$, where I_d is the $d\times d$ identity matrix [6]. This leads to the expression

$$\hat{f}(\mathbf{x};h)=\frac{1}{nh^d}\sum_{i=1}^{n}K\left(h^{-1}(\mathbf{x}-\mathbf{X}_i)\right) \qquad (1.2)$$

To use the parameterization $H=h^2 I_d$ effectively, the components of the data vector should be commensurate. This can be achieved by using appropriate transformation in the data set [6], [20], [22]. This transformation involves either pre-scaling each axis (that is , normalize to unit variance, for instance) or pre- whitening the data (that is, linearly transform to have unit covariance matrix). A detailed study of this can be found in [8]. The transformation guarantees the use of the form involving single bandwidth as in (1.2).

Many of the studies in density estimation have been centred on the univariate kernel density estimators [20]. However, this paper focuses on the multivariate settings with emphasis on the efficiency of some classical product

Manuscript received October 02, 2013, revised July 17, 2014.
*F. O. Oyegue is with the Department of Mathematics, Faculty of Physical Sciences, University of Benin, Benin City, Edo State, Nigeria (fooyegue@yahoo.com).
S. M. Ogbonmwan is with the Department of Mathematics, Faculty of Physical Sciences, University of Benin, Benin City, Edo State, Nigeria (ogbonmwasmaltra@yahoo.co.uk).

(multivariate) kernels. The concept of efficiency is used in kernel density estimation to analyse the effect of second-order multivariate kernels so that an appropriate kernel can be chosen.

The basic motivation for considering (1.2) is that it enables one to obtain closed form expressions for the optimal bandwidth and the asymptotic mean integrated squared error (AMISE). Thus we derive the the generalized expression for the efficiency of second-order multivariate symmetric kernel. Throughout this paper, $\int$ is the shorthand for $\int_{R^d}$. The global accuracy used in measuring (1.2) is the mean integrated square error (MISE). The expression for the MISE is

$$MISE(h) = E\int \left(\hat{f}(\mathbf{x}) - f(\mathbf{x})\right)^2 d\mathbf{x} \qquad (1.3)$$

Thus, from [21], the expression (1.3) can be written as a sum of integrated square bias and integrated variance of $\hat{f}_h(\mathbf{x})$. That is,

$$MISE\left\{\hat{f}_H(\mathbf{x})\right\} = \int \left(E\left\{\hat{f}_H(\mathbf{x})\right\} - f(\mathbf{x})\right)^2 d\mathbf{x} + \int Var\left\{\hat{f}_H(\mathbf{x})\right\} d\mathbf{x} \qquad (1.4)$$

The concept of efficiency for univariate kernels was popularized by [20], and this was followed by the work [21] who gave an insight into the efficiency of the second-order multivariate kernels. [21] approach was based on taking the ratio of the spherically symmetric kernel relative to the product kernel. Hence, we develop a method that is different from the approach adopted by [21], even though our method is motivated by the work [20] and [21].

The remainder of this paper is as follows. In section 2, we cover the necessary background materials on the asymptotic mean integrated square error (AMISE). In section 3, the generalized expression for the efficiency of second – order multivariate kernels is derived. In section 4, we compare the efficiencies of multivariate kernels for the cases $d = 1, 2, 3, 4, and\ 5$.

II. THE AMISE FOR THE MULTIVARIATE KERNEL DENSITY ESTIMATOR

The asymptotic mean integrated square error (AMISE) is one of the most important parts in bandwidth selection. By using symmetric kernel function, the AMISE and the optimal bandwidth for the multivariate kernel density estimator are derived. The kernel determines the slope of the estimator, while the amount of smoothing is determine by the bandwidth h. In particular $\hat{f}_H(\mathbf{x})$ as define in (1.2) is a density function provided $K(\mathbf{w}) \geq 0$ and $\int K(\mathbf{w})d\mathbf{w} = 1$,

where, $\mathbf{w} = H^{-1}(\mathbf{x}-\mathbf{y})$ [11]

In the case $d \geq 1$, the most often used choice is a density function, which6is symmetric about zero, and such that

$$\begin{cases} \int \mathbf{w}K(\mathbf{w})d\mathbf{w} = O_d \\ \int \mathbf{w}\mathbf{w}^T K(\mathbf{w})d\mathbf{w} = \mu_2\ I_d \end{cases}, \qquad (1.5)$$

[19].

The usual criterion for the optimal bandwidth is the asymptotic version of the MISE in (1.3) [5], [9], [12], [14], [20]. To find the AMISE, one needs to find the bias and variance of $\hat{f}_h(\mathbf{x})$. $E\left(\hat{f}_H(\mathbf{x})\right)$ can be evaluated by using Taylor series expansion on

$$E\left(\hat{f}_H(\mathbf{x})\right) = \frac{1}{|H|}\int K_H\ (\mathbf{x}-\mathbf{y}) f(\mathbf{y})d\mathbf{y}$$

$$= \int K(\mathbf{w})\ (f(\mathbf{x}) - tr\ (H^{\frac{1}{2}}\ Df\ (\mathbf{x})\mathbf{w}^T$$

$$+ tr\ [H^{\frac{1}{2}}\ D^2 f(\mathbf{x}) f(\mathbf{x})\ H^{\frac{1}{2}}\ \mathbf{w}\mathbf{w}^T]))d\mathbf{w} \qquad (1.6)$$

to second order [21].

Now imposing the conditions (1.5) and $\int k(\mathbf{w})\ d\mathbf{w} = 1$ on (1.6), results in

$$E\left(\hat{f}_H\ (\mathbf{x})\right) = f(\mathbf{x}) + \frac{1}{2!}\mu_2\ (K)\ tr\ \left(H^{\frac{1}{2}} D^2\ f(\mathbf{x})\ H^{\frac{1}{2}}\right)$$

Hence the bias term becomes

$$Bias\left(\hat{f}_H(\mathbf{x})\right) = \frac{1}{2!}\mu_2\,(K)\ tr\ \left(H^{\frac{1}{2}} D^2\ \ f(\mathbf{x})\ H^{\frac{1}{2}}\right)$$ and

the asymptotic integrated square bias (AISB) becomes

$$AISB\left(\hat{f}_H(\mathbf{x})\right) \cong \int bias^2\ \left(\hat{f}_H\ (\mathbf{x})\right) d\mathbf{x}$$

$$\cong \frac{1}{(2!)^2}\mu_2\ (k)^2 \int tr\ \left[H^{\frac{1}{2}}\ D^2 f(\mathbf{x})\ H^{\frac{1}{2}}\right]^2 d\mathbf{x}. \quad (1.7)$$

The variance term is [21]. That is,

$Var\ \left(\hat{f}_H\ (\mathbf{x})\right) = \frac{f(\mathbf{x})}{n|H|^{\frac{1}{2}}}\int k(\mathbf{w})^2\ d\mathbf{w}$ and hence the asymptotic integrated variance (AIV) becomes:

$$AIV\ \left(\hat{f}_H\ (\mathbf{x})\right) = \frac{R(K)}{n|H|^{\frac{1}{2}}} \qquad (1.8)$$

Combing (1.7) and (1.8) yield

$$AMISE\ \left\{\hat{f}_H(\mathbf{x})\right\} = \frac{R(k)}{n|H|^{\frac{1}{2}}} +$$

$$\frac{1}{(2!)^2}\mu_2\ (K)^2 \int tr\ \left(H^{\frac{1}{2}} D^2\ f(\mathbf{x})\ H^{\frac{1}{2}}\right)^2\ d\mathbf{x}$$

and since $H = h^2\ I_d$, it results to

$$AMISE\left(\hat{f}_h(\mathbf{x})\right) = \frac{R(k)}{nh^d} + \frac{1}{(2!)^2}\mu_2(k)^2 h^4 \int \left(\nabla^2 f(\mathbf{x})\right)^2 d\mathbf{x} \qquad (1.9)$$

Minimization of (1.9) with respect to h leads to the formula for the optimal bandwidth in the following form

$$h_{opt} = \left\{\frac{dn^{-1} R(K)}{\mu_2(K) R\left(\nabla^2 f(\mathbf{x})\right)}\right\}^{\frac{1}{d+4}} \qquad (2.0)$$

where $R\left(\nabla^2 f\right) = \int \left(\nabla^2 f(\mathbf{x})\right)^2 d\mathbf{x} < \infty$, $D^2 f(\mathbf{x}) = \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i} \mathbf{x}$

Putting (2.0) into (1.9), the minimum AMISE is obtained as;

$$AMISE\left(\hat{f}_h(\mathbf{x})\right) = \left(\frac{d+4}{4d}\right) \times \left\{\begin{matrix}\mu_2(K)^{2d} \left(dR(K)^4\right) \times \\ \left(\int \left(\nabla^2 f(\mathbf{x})\right)^2 d\mathbf{x}\right)^d n^{-4}\end{matrix}\right\}^{\frac{1}{d+4}} \qquad (2.1)$$

Equation (2.0) is a closed form solution for the bandwidth vector which minimizes the expression for the AMISE in (2.1). Moreover , the optimal bandwidth is of order $n^{-\frac{1}{d+4}}$ and the optimal AMISE is of order $n^{-\frac{4}{d+4}}$.

III. EFFICIENCY FOR THE SECOND ORDER MULTIVARIATE KERNELS

In this section, the AMISE expression so derived is used to develop the generalized expression for the efficiency of second order multivariate kernels.

One way of obtaining the multivariate forms of any unvariate kernel is by using the product kernel method which is given by [21] as;

$$K^p(\mathbf{x}) = \prod_{i=1}^{d} K(x_i) \qquad (3.1)$$

Where, $K(x)$ is the univariate symmetric kernel.

The efficiency of the univariate symmetric kernel defined by [20] is

$$Eff(K) = \left\{\frac{C(K_e)}{C(K)}\right\}^{\frac{5}{4}} \qquad (3.2)$$

where $C(K) = \left\{\int x_1^2 K(\mathbf{x}) d\mathbf{x}\right\}^{\frac{2}{5}} \left\{\int K(\mathbf{x})^2 d\mathbf{x}\right\}^{\frac{4}{5}}$ is any given kernel constant under discussion and $C(K_e) = \left\{\int x_1^2 K_e(\mathbf{x}) d\mathbf{x}\right\}^{\frac{2}{5}} \left\{\int K_e(\mathbf{x})^2 d\mathbf{x}\right\}^{\frac{4}{5}}$ is the Epanechnikov kernel constant.

By drawing inspiration from equation (3.1), the general expression for the efficiency of multivariate kernels based on the product kernel approach is now defined as

$$Eff\left(K^p(\mathbf{x})\right) = \left\{\frac{C_d^2\left(K_e^p\right)}{C_d^2\left(K^p\right)}\right\}^{\frac{d+4}{4}} \qquad (3.3)$$

where

$C_d^2\left(K^p\right) = \left\{\int x_1^2 K^p(\mathbf{x}) d\mathbf{x}\right\}^{\frac{2d}{d+4}} \left\{\int K^p(\mathbf{x})^2 d\mathbf{x}\right\}^{\frac{4}{d+4}}$ is the d-dimensional product form of any given second order kernel constant and $C_d^2\left(K_e^p\right) = \left\{\int x_1^2 K_e^p(\mathbf{x}) d\mathbf{x}\right\}^{\frac{2d}{d+4}} \left\{\int K_e^p(\mathbf{x})^2 d\mathbf{x}\right\}^{\frac{4}{d+4}}$ is d - dimensional product form of the Epanechnikov kernel constant.

Theorem1. If equation (3.3) holds, then the efficiency for the second – order d -dimensional kernel is

$$Eff\left\{K^p(\mathbf{x})\right\} = \left(\frac{3}{5\sqrt{5}}\right)^d \times \left\{\int x_1^2 K^p(\mathbf{x}) d\mathbf{x}\right\}^{-\frac{d}{2}} \left\{\int K^p(\mathbf{x})^2 d\mathbf{x}\right\}^{-1}$$

Proof:

The univariate Epanechnikov kernel as defined in [20] is

$$K(x) = \frac{3}{4\sqrt{5}}\left(1 - \frac{x^2}{5}\right), -\sqrt{5} \le x \le \sqrt{5}$$

Hence, the multivariate version using (3.1) is

$$K_e^p(\mathbf{x}) = \prod_{i=1}^{d}\left(\frac{3(5 - x_i)^2}{20\sqrt{5}}\right)$$

From (2.1), set

$$C_d^2(K) = \mu_2(K)^{\frac{2d}{d+4}} R(K)^{\frac{4}{d+4}} = \left\{\int x_1^2 K(\mathbf{x}) d\mathbf{x}\right\}^{\frac{2d}{d+4}} \left\{\int K(\mathbf{x})^2 d\mathbf{x}\right\}^{\frac{4}{d+4}} \quad (3.4)$$

Re-write equation (3.4) to reflect (3.3). That is

$$C_d^2\left(K_e^p\right) = \left\{\int x_1^2 K_e^p(\mathbf{x}) d\mathbf{x}\right\}^{\frac{2d}{d+4}} \left\{\int K_e^p(\mathbf{x})^2 \right\}^{\frac{4}{d+4}}$$

and

$$C_d^2\left(K^p\right) = \left\{\int x_1^2 K^p(\mathbf{x}) \right\}^{\frac{2d}{d+4}} \left\{\int K^p(\mathbf{x})^2 d\mathbf{x}\right\}^{\frac{4}{d+4}}$$

Thus;

$$\mu_2\left(K_e^p\right) = \int x_1^2 K_e^p(\mathbf{x}) d\mathbf{x}$$

$$= \int x_1^2 \left(\prod_{i=1}^{d} \frac{3\left(5 - x_i^2\right)}{20\sqrt{5}}\right) d\mathbf{x} = 1 \quad \text{and}$$

ISBN: 978-988-19253-7-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$R\left(K_e^p\right) = \int K_e^p(\mathbf{x})^2 \; d\mathbf{x}$$

$$= \int \left(\prod_{i=1}^{d} \frac{3\left(5 - x_i^2\right)}{20\sqrt{5}} \right)^2 d\mathbf{x} = \left(\frac{3}{5\sqrt{5}} \right)^d$$

Putting the values of $\mu_2\left(K_e^p\right)$ *and* $R\left(K_e^p\right)$ into (3.3) yields

$$Eff\ \left\{R^p(\mathbf{x})\right\} = \left\{ \frac{\left[\left(\frac{3}{5\sqrt{5}}\right)^d\right]^{\frac{4}{d+4}}}{\left[\left(\int x_1^2\ K^p(\mathbf{x})d\mathbf{x}\right)^{2d}\left(\int K^p(\mathbf{x})^2\ d\mathbf{x}\right)^4\right]^{\frac{1}{d+4}}} \right\}$$

Hence,

$$Eff\ [K(\mathbf{x})] = \left(\frac{3}{5\sqrt{5}}\right)^d \times$$

$$\left(\int x_1^2\ K^p(\mathbf{x})\ d\mathbf{x}\right)^{-\frac{d}{2}} \left(\int K^p(\mathbf{x})^2\ d\mathbf{x}\right)^{-1} \qquad (3.5)$$

From the equation (3.5) which is the generalized expression for the second order multivariate kernel, the efficiencies of some *d*-dimensional (*for* $d = 1,2,3,4,5$) kernel derived from some univariate kernels (i.e the uniform, biweight, triweight, and Gaussian); using mathematica 6.0 platform, are obtained; and their graphs, using excel, are shown in Fig.1
.

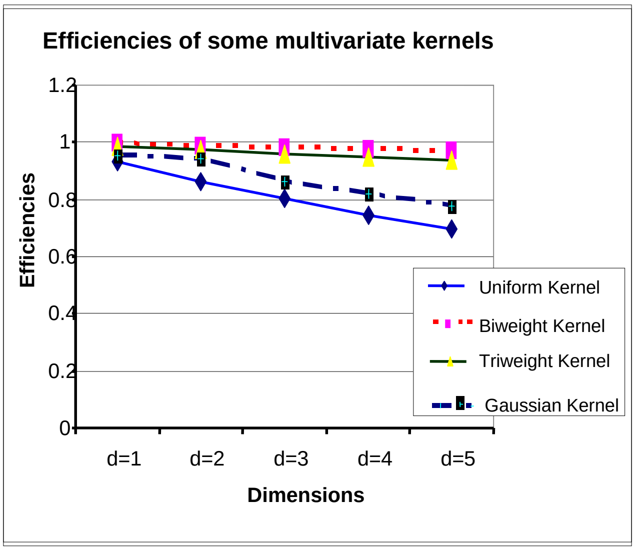

Fig.1.Efficiency of some multivariate (i.e.*d* = 1, 2, 3. 4, 5) kernels.

Examining Fig.1 for higher dimensional kernels, that is, for example, *d* = 2, it is observed that relative to the Epanechnikov kernel, there is a 14% loss in efficiency for the uniform kernel, the Gaussian lost about 10% in efficiency; the biweight and the triwegtht shed about 1% and 3% respectively in efficiency. For *d* = 3, 4, 5, the uniform kernel lost about 20%, 25%, 31%, respectively in efficiency; the biweight and the triweight lost approximately 2%, 2%, 3% and 4%, 5%, 6% respectively in efficiency. There is a loss of about 14%, 18%, and 22% when *d* = 3, 4, and 5 respectively in the case of the Gaussian kernel. Furthermore, a comparison of the dimensions of the four beta kernels (the Gaussian, the uniform, the biweight, and the triweight), shows that the biweight and the triweight kernels give relatively better efficiencies than the uniform and the Gaussian kernels. This is visible in Fig.1, where there is a slight drop in the efficiencies of both the biweight and the triweight kernels as their dimension increases with the biweight having an edge over the triweight kernel. From the same Fig.1, it is observed that there is a sharp fall in the efficiencies of both the uniform and the Gaussian kernels as their dimension increases with the uniform kernel becoming appreciably worse. This clearly shows that in contrast to the biweight and the triweight kernels, the uniform kernel and the Gaussian kernel are highly inefficient with the efficiency loss increasing as the dimension increases. Although, we observed that for the various kernels considered, their efficiencies decrease as the dimension increases; the calculations suggest that the biweight kernel and the triweight kernel are good choices of density estimators.
In all, the implication of this is that, for example, in the case of $d = 2$ for the Gaussian kernel, the minimum $AMISE(f)$ obtained using the Epanechnikov kernel with a sample size of $n = 90$ is approximately equal to the minimum $AMISE(f)$ obtained using the Gaussian kernel $n = 100$.

IV. CONCLUSION

In this paper, a new computational approach has been developed for the efficiency of multivariate product kernels. The Epanechnikov kernel was used as a theoretical underpinning for deriving the efficiency formula. The new efficiency formula was experimented with four of the beta kernels, viz.: the Gaussian, the uniform, the biweight, and the triweight kernels. Findings revealed that the biweight and the triweight kernels have relatively high efficiency values. By this, we infer that they are better density estimators than the Gaussian and the uniform kernels form of the multivariate product kernels. Nevertheless, it is premature to conclude that the biweight and the triweight kernels are the most suitable multivariate product kernels. This is because the spherical aspects of the multivariate kernels have not been considered. We therefore suggest the development of a theoretical framework for the efficiency of multivariate kernels using the spherical methods as a grey area for future research.

ISBN: 978-988-19253-7-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

REFERENCES

[1] B. A. Akaiki, “An approximation to the density function,”. *Annals of Institute of Statistical Mathematics*, Vol.6, pp. 127-132, June 1954.

[2] A.W. Bowman and A. Azzalini, *Applied smoothing technique for data analysis: The kernel approach with S-plus illustration*. Oxford: Oxford University Press, 1997, pp. 1-2.

[3] T. Brox, B. Rosenhahn, U. Kersting,, and D. Cremers, “Nonparametric density estimation for human pose tracking,” *In Lecture Notes in Computer Science*: *Pattern Recognition*, Vol. 4174, pp. 546-555, September, 2006.

[4] T. Cacoullos, “Estimation of a multivariate density,” *Annals of Institute of Statistical Mathematics*, Vol. 18, pp. 178-189, December 1966..

[5] L. Devroye and L. Gjorfi, *Nonparametric density estimation: The L1 View*. New York: Wiley, 1984, pp. 152-156.

[6] T. Duong and M. L. Hazeltom, “Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimator,” *Journal of Multivariate Analysis,* Vol. 93, pp. 417-433, April 2005.

[7] E. Fix and J.L. Hodges, “ Discriminatory analysis- nonparametric discrimination: Consistency properties,” USAF School of Aviation Medicine, Randolf Field, Texas, Report Number 4, Project Number 21-49-004, February 1951.

[8] K. Fukunaga, *Introduction to statistical pattern recognition*. New York: Academic Press, 1990, pp. 46-49.

[9] P. Halll and J.S. Marron, “Choice of kernel order in density estimation,” *Annals of Statistics*, Vol. 16, pp. 161-173, November 1987.

[10] W. Hardle, M. Muller, S. Sperlich, and A. Werwatz. *Nonparamtric and semiparametric models*. Berlin: Springer-Verlag, 2004, pp. 21-46.

[11] N. Howe, M. Leventon, and W. Freeman, “Baysian reconstruction of 3D human motion from single-camera video,” *Proceedings of NIPS International Conference on on Advances in Neural Information Processing Systems*, Vol. 12, pp. 820-826, November 1999.

[12] J. Jarnicka, “ Multivariate kernel density estimation with a parametric support,” *Opuscula Mathematica,* Vol. 29, No.1, pp. 41-55, 2009..

[13] C.G. Lambert, S.E., Harrington, C.E. Harvey and A. Glodjo, “Efficient online non-parametric kernel density estimation,” *Algorithmica*, Vol. 25, pp. 37-57, May 1999.

[14] J.S. Marron and M.P. Wand , “Exact mean integrated squared error,” *Annals of Statistics,* Vol. 20, No. 2, pp. 712-736, June 1992.

[15] P.D. Michailidis and K.G. Margritis, “Accelerating kernel density estimation on the GPU using the CUDA framework,” *Applied Mathematical* Sciences, Vol. 7, No. 30, pp. 1447-1476, January 2013,.

[16] E. Parzen, “On the estimation of a probability density function and the mode,” *Annals of Mathematical* Statistics, Vol. 33, No. 3, pp. 1065-1076, September 1962.

[17] M. Rosenblatt, “ Remarks on some nonparametric estimate of a density function,” *Annals of Mathematical Statistics,* Vol. 27, No. 3, pp. September 1956.

[18] B. Rosehahn, T. Brox, and J. Weikert, “Three dimensional shape knowledge for joint image segmentation and pose tracking,” *International Journal of Computer Vision*, Vol. 73, No. 3, pp. 243-262, July 2007.

[19] D. W. Scott, *Multivariate Density Estimation: Theory, Practice and Visualization*. New York: *Wiley*, 1992, pp. 153.

[20] B.W. Silverman, *Density estimation for statistics and data analysis*. London: Chapman and Hall, 1986, pp. 41-85.

[21] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London: Chapman and Hall, 1995, pp. 91-109.

[22] T. J. Wu, C. F. Chen, and H. Y. A. Chen, “Variable bandwidth selector in multivariate kernel density estimation,” *Ststistics and Probability Letters*, Vol. 77, No. 4, pp. 462-467, February 2007.

[23] Y. Zheng , J. Jestes, J.M., Phillips, and F. Li, “Quality and efficiency in kernel density estimation for large data,” *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 433-444, June 2013.