International Journal of Architectural, Civil and Construction Sciences
ISSN: 2415-1734
Vol:9, No:5, 2015

Mean Shift-based Preprocessing Methodology for Improved 3D Buildings Reconstruction

Nikolaos Vassilas, Theocharis Tsenoglou, Djamchid Ghazanfarpour

***Abstract*—**In this work, we explore the capability of the mean shift algorithm as a powerful preprocessing tool for improving the quality of spatial data, acquired from airborne scanners, from densely built urban areas. On one hand, high resolution image data corrupted by noise caused by lossy compression techniques are appropriately smoothed while at the same time preserving the optical edges and, on the other, low resolution LiDAR data in the form of normalized Digital Surface Map (nDSM) is upsampled through the joint mean shift algorithm. Experiments on both the edge-preserving smoothing and upsampling capabilities using synthetic RGB-*z* data show that the mean shift algorithm is superior to bilateral filtering as well as to other classical smoothing and upsampling algorithms. Application of the proposed methodology for 3D reconstruction of buildings of a pilot region of Athens, Greece results in a significant visual improvement of the 3D building block model.

***Keywords*—**3D buildings reconstruction, data fusion, data upsampling, mean shift.

I. Introduction

IN the recent years, methods like bilateral and mean shift filtering are becoming increasingly popular at least in the fields of image processing and computer vision due to their edge preserving smoothing capabilities [1]-[4]. Moreover, simple modifications in the above filtering methods leading to the so called joint bilateral and joint mean shift filtering permit fusion of spatial data from different sources and of different resolutions. The result of data fusion is to obtain improved interpolation weights on the low resolution data sources which is reflected to improved - qualitatively and quantitatively - upsampled data [5].

Joint bilateral filtering has been used in various applications, such as, in digital flash photography whereby flash images are used to increase sharpness of non-flash ambient images [6], [7], in improving resolution of depth maps obtained from pairs of stereo images [5], and 3D-mesh smoothing [8], just to name a few.

While bilateral filtering has become quite popular filtering method in the recent years, we have the feeling that mean shift has not yet found its place among the most powerful choices for image clustering, segmentation, edge preserving smoothing and upsampling. Few works have performed comparative experiments with other popular methods in order to assess the qualities and competitiveness of the algorithm in the various application domains. In this respect, we mention [4] which provides qualitative comparisons on the clustering capabilities of the following algorithms: iterative bilateral filtering, nonlinear diffusion, restricted mean shift (i.e. with fixed spatial component) and the joint spatial/range mean shift algorithm. Even worse, there has not been any work that we are aware of, on the use of joint mean shift filtering for data upsampling.

In this work, we first establish the power of the mean shift algorithm for image denoising and data upsampling through comparative experiments and then use it for improving the 3D models of buildings when the elevation data (e.g. the nDSM) are of much lower spatial resolution than the corresponding optical data.

Section II provides the necessary background and presents our proposed version of the algorithm. Section III compares several smoothing trechniques in the field of JPEG artifact removal. Section IV compares popular upsampling methods. Section V employs the proposed mean shift-based preprocessing technique for 3D building modeling and, finally, Section VI presents the conclusions and future extensions of this work.

II. Joint Mean Shift Filtering

In this Section we present a novel data preprocessing method appropriate for 3D building reconstruction in order to increase spatial resolution of elevation data. In the proposed method, low resolution elevation data (e.g. produced by airborne Light Detection and Ranging (LiDAR) sensors) are fused with high resolution RGB color aerial orthophotographs through the mean shift algorithm. The result is improved quality for both the optical and the upsampled elevation data. Following a short presentation of the mean shift algorithm, the section concludes with the presentation of the mean shift-based data fusion and upsampling method.

A. The Mean Shift Algorithm

Mean shift is a non-parametric, iterative algorithm for finding the local maxima in a density function. Although it was first proposed by Fukunaga and Hostetler about 40 years ago [9], re-examined by Cheng [10] and, later, by Weijer and Boomgaard [11], only recently it has been introduced by Comaniciu and Meer [2] to low-level vision problems, such as edge preserving smoothing, segmentation and clustering.

This method makes no prior assumptions about the form of the density function or, in case of clustering, about the number

N. Vassilas is with the Department of Informatics, Technological Educational Institute (T.E.I.) of Athens, Athens, Greece (phone: ++30 210 5385827; fax: ++30 210 5910975; e-mail:nvas@teiath.gr).

Th. Tsenoglou is with the Department of Informatics, T.E.I. of Athens, Athens, Greece (e-mail: ttheoharis@hotmail.com).

D. Ghazanfarpour is with the Department of Mathematics & Informatics, XLIM (UMR CNRS 7252), University of Limoges, Limoges, France (e-mail: djamchid.ghazanfarpour@unilim.fr).

of clusters. Using a kernel-based approximation of the underlying density, the algorithm employs fixed point iteration to solve the nonlinear maximization problem of locating the modes (i.e. the maxima) of the density. Specifically, the iterations are initialized with a point in feature space. Then each iteration consists of two steps:

In the first step, it computes the point of highest density in a neighborhood of the current estimate by evaluating the weighted average of the feature values in this neighborhood. The weights for computing the average and the size of the neighborhood are chosen in advance. They are determined by the selection of the kernel function and its bandwidth, h.

In the second step, the mode estimate is updated by moving towards the point of highest concentration.

These two steps are repeated until there is no further modification in the values of the mode estimates. The speed of convergence and the accuracy of the final value depend on the kernel chosen and the size of the neighborhood. The algorithm uses one kernel for each different type of feature that constitutes the dimensions of the data (range) space. For one type of feature and one kernel of bandwidth h, the update formula for the mean shift is

$$\mathbf{x}(n+1)=\frac{\sum_i \mathbf{z}_i\, k(\| \mathbf{x}(n)-\mathbf{z}_i \|^2/h^2)}{\sum_i k(\| \mathbf{x}(n)-\mathbf{z}_i \|^2/h^2)} \tag{1}$$

where n is the iteration index, $\{\mathbf{z}_i \mid i = 1, \ldots, N\}$ is the initial data set, $k(\ )$ is the interpolating kernel, often selected to be gaussian or box shaped and $\mathbf{x}(n)$ is the point trajectory in feature space.

B.Joint Mean Shift for Data Fusion and Upsampling

Because our LiDAR data are of low quality, improving their resolution and sharpness requires using additional information from other sources.

For this purpose, in our approach, we fuse the elevation information with a high resolution orthophoto color image of the same region through the joint mean shift algorithm in a similar way to the joint (or cross) bilateral filtering as it has been used in the field of digital photography for denoising ambient images [6], [7] and for data upsampling [5]. The implicit assumption is that the optical data can provide the necessary information about the significant edges. The high detail content of the color image will be a guide for improving the quality of the elevation image. But the optical data also contain a great amount of unnecessary noisy edges. So, the problem is, generally, twofold:

- to smooth small color variations in areas of small elevation variations and smooth out height variations, due to noise, in areas of relatively flat color content.
- to preserve the significant optical edges.

To achieve this we use a variant of the mean shift algorithm that operates jointly in the spatial and range domains whereby the range domain comprises of both optical (RGB) and elevation data. During the iterative process and through the coupling of the data the aim is to fuse the significant edges while smoothing the noise.

The proposed methodology comprises the following two stages. First, an initial upsampling using the typical nearest neighbor interpolation technique is performed on the low resolution elevation data (nDSM) to increase its size to the size of the color image.

Next, in order to improve the quality of the result and eliminate the staircase effects of nearest neighbor upsampling near elevation discontinuities, we perform a mean shift-based discontinuity preserving smoothing on the combined spatial, optical and elevation data. For this purpose, each pixel i is represented by a feature vector $\mathbf{z}_i$ that includes the three color components, $\mathbf{z}_i^{\mathrm{c}}$, its elevation value z_i^{e} and its spatial coordinates $\mathbf{z}_i^{\mathrm{s}}$. For simplicity, we have chosen all three kernels to be gaussian (although uniform kernels constitute, also, good alternatives) with the respective bandwidths, h_{s}, h_{c} and h_{e}, of the kernels being chosen through experimentation. In this case, the formula for the joint feature vector update becomes:

$$\begin{bmatrix}\mathbf{x}_i^{\mathrm{s}}(n+1)\\ \mathbf{x}_i^{\mathrm{c}}(n+1)\\ x_i^{\mathrm{e}}(n+1)\end{bmatrix}=\frac{1}{K_j(n)}\sum_i \begin{bmatrix}\mathbf{z}_i^{\mathrm{s}}(n)\\ \mathbf{z}_i^{\mathrm{c}}(n)\\ z_i^{\mathrm{e}}(n)\end{bmatrix}\exp\left(-\frac{\| \mathbf{x}_j^{\mathrm{c}}(n)-\mathbf{z}_i^{\mathrm{c}} \|^2}{h_{\mathrm{c}}^2}\right) \cdot \exp\left(-\frac{\| x_j^{\mathrm{e}}(n)-z_i^{\mathrm{e}} \|^2}{h_{\mathrm{e}}^2}\right)\exp\left(-\frac{\| \mathbf{x}_j^{\mathrm{s}}(n)-\mathbf{z}_i^{\mathrm{s}} \|^2}{h_{\mathrm{s}}^2}\right) \tag{2}$$

where n is the iteration index, i, j are data indices and the normalization constant $K_j(n)$ is computed as

$$K_j(n)=\sum_i \exp\left(-\frac{\| \mathbf{x}_j^{\mathrm{c}}(n)-\mathbf{z}_i^{\mathrm{c}} \|^2}{h_{\mathrm{c}}^2}\right)\exp\left(-\frac{\| x_j^{\mathrm{e}}(n)-z_i^{\mathrm{e}} \|^2}{h_{\mathrm{e}}^2}\right) \cdot \exp\left(-\frac{\| \mathbf{x}_j^{\mathrm{s}}(n)-\mathbf{z}_i^{\mathrm{s}} \|^2}{h_{\mathrm{s}}^2}\right) \tag{3}$$

The summations are over all pixels i in a spatial neighborhood of the current pixel j. Equation (2) is computed iteratively for every pixel.

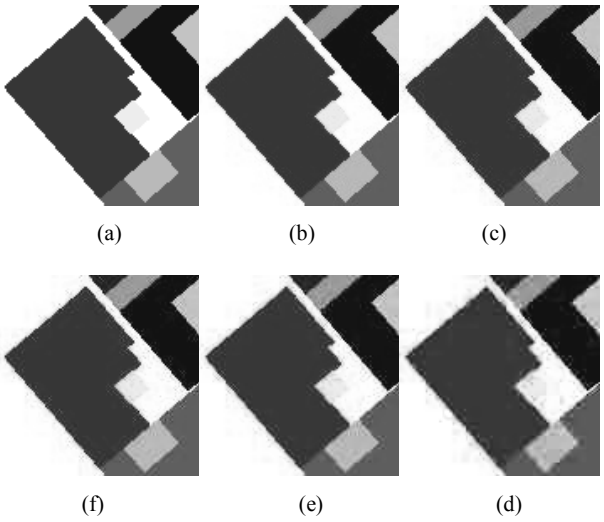
Equation (2) clearly indicates the interdependency, during the updating step, of the color and the elevation values. For example, if a pixel has color value $\mathbf{x}_j^{\mathrm{c}}(n)$ that differs considerably from some of the neighbors $\mathbf{z}_i^{\mathrm{c}}$ then these neighbors will not contribute in the computation of either the color or the elevation mean update of the pixel nor in the new spatial coordinates. The same is true of the influence of the elevation differences in computing the updates. This has the effect that if for a pixel there is a large discrepancy in one space with some of its neighbors then this is mirrored in the other space.

The results of the processing are, on one hand, an edge preserving smoothing of the RGB image, and on the other, hand; elevation image with much straighter height discontinuities. Color variations, due to noise, in flat surfaces of the RGB image become more homogeneous without losing the significant optical edges. At the same time, the elevation

image gains significantly in detail and sharpness with the different elevation surfaces becoming much better discriminated. It should be noted that edges due to shadows in the color image do not appear in the resulting elevation image if they do not correspond to significant elevation variations.

III. Comparisons on Smoothing Techniques

In this section, mean shift is compared to other smoothing algorithms for the removal of JPEG artifacts in synthetic images. In particular, we generated 5 JPEG images at various compression rates that correspond to JPEG quality levels of 90, 70, 50, 30 and 10 of a prototype RGB image of 120x120 pixels, i.e. of size 43,200 bytes. Fig. 1 (a) shows the prototype image and Figs. 1 (b)-(f) show the above compressed JPEG images. The JPEG lossless compression file of the original prototype image is of 6704 bytes while the corresponding lossy compression sizes of Figs. 1 (b)-(f) are 5727, 3649, 2884, 2289 and 1535 bytes, respectively. The lossless compression ratio of Fig. 1 (a) is 6.4:1 while the corresponding lossy compression ratios for Figs. 1 (b)-(f) are 7.6:1, 11.8:1, 15.0:1, 18.9:1 and 28.1: 1.



(a) (b) (c)

(f) (e) (d)

Fig. 1 (a) The prototype image, (b)-(f) the compressed images that correspond to JPEG quality of 90, 70, 50, 30 and 10 respectively

The severity of the JPEG artifacts of the above images is assessed through the root mean square error (RMSE) of the color deviations (in gray levels) from the ones of the original image. The RMSEs of the above 5 cases are 12.18, 14.81, 17.28, 19.82 and 24.06 gray levels, respectively.

The experiments have been performed in both the RGB and L*a*b color spaces and the resulting RMSEs for uniform kernels and optimal h_s and h_c parameter selection are shown in Table I. From Table I, it is evident that the RGB color space gave better noise removal than the L*a*b color space. Consequently, we decided to use the RGB color space in all subsequent experiments.

TABLE I
Comparison of RGB and L*a*b Color Spaces

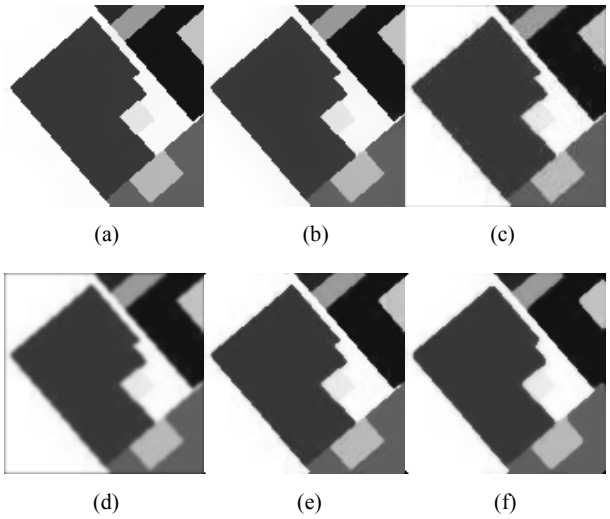
JPEG Quality	Initial RMSE	RGB: RMSE for h_S =11, h_C = 0.4	L*a*b: RMSE for h_S = 7, h_C = 0.25
90	12.18	3.79	5.47
70	14.81	4.78	7.07
50	17.28	6.59	8.92
30	19.82	8.94	13.15
10	24.06	15.29	17.88

Next, we performed comparisons of mean shift smoothing with bilateral filtering, Gaussian smoothing and median filtering. Both mean shift and bilateral filtering used uniform kernels with h_s = 11 and h_c = 0.4. Two gaussian filters with 3x3 and 5x5 templates and σ = 0.5 and 1.0 respectively have been used for comparisons along with 3x3 and 5x5 median filters. The results are tabulated in the form of RMSEs in Table II. It is apparent from these results that the edge preserving smoothing achieved by the first two algorithms is the reason for the significantly better JPEG artifact removal.

TABLE II
Comparison of Smoothing Algorithms for JPEG Artifact Removal

JPEG Quality	Mean shift	Bilateral filtering	Gaussian Filtering		Median Filtering	
			3x3 σ = 0.5	5x5 σ = 1	3x3	5x5
90	3.79	6.17	14.79	23.06	10.76	14.08
70	4.78	7.82	16.65	24.06	13.02	15.56
50	6.59	9.62	18.54	24.95	15.56	17.29
30	8.94	11.99	20.74	25.93	18.47	19.47
10	15.29	17.45	24.39	28.08	22.87	23.16

The smoothed images for the quality level of 50 for the 6 methods used in the above comparisons are shown in Fig. 2.



(a) (b) (c)

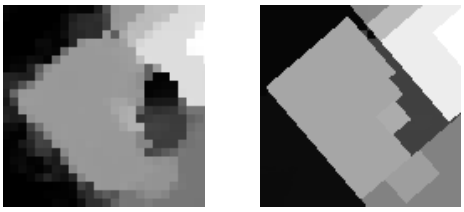
(d) (e) (f)

Fig. 2 The smoothed result of Fig. 1 (d) using: (a) mean shift, (b) bilateral filtering, (c) Gaussian 3x3 filter with σ = 0.5, (d) Gaussian 5x5 filter with σ = 1, (e) 3x3 median filter and (f) 5x5 median filter

IV. Comparisons on Upsampling Techniques

In the second set of comparative experiments, we fused the high resolution optical data with low resolution elevation data

to improve the corresponding 3D representation. In fact, the improved 3D model is achieved through a better interpolation method guided by the mean shift algorithm. At first, we used the joint spatial/optical/elevation update version of mean shift algorithm (the range corresponded to the RGB values and the elevation of each pixel) to improve the elevation (LiDAR) data shown in Fig. 3 (a) with the prototype (segmented) data of Fig. 1 (a). Also shown, in Fig. 3 (b), is the resulted elevation image for $h_s = 35$, $h_c = 0.05$ and $h_e = 0.2$. The initial and final 3D representations of the elevation data are shown in Figs. 4 (a) and (b) respectively. For qualitative comparisons, in Fig. 4 (c) the prototype 3D model is shown.



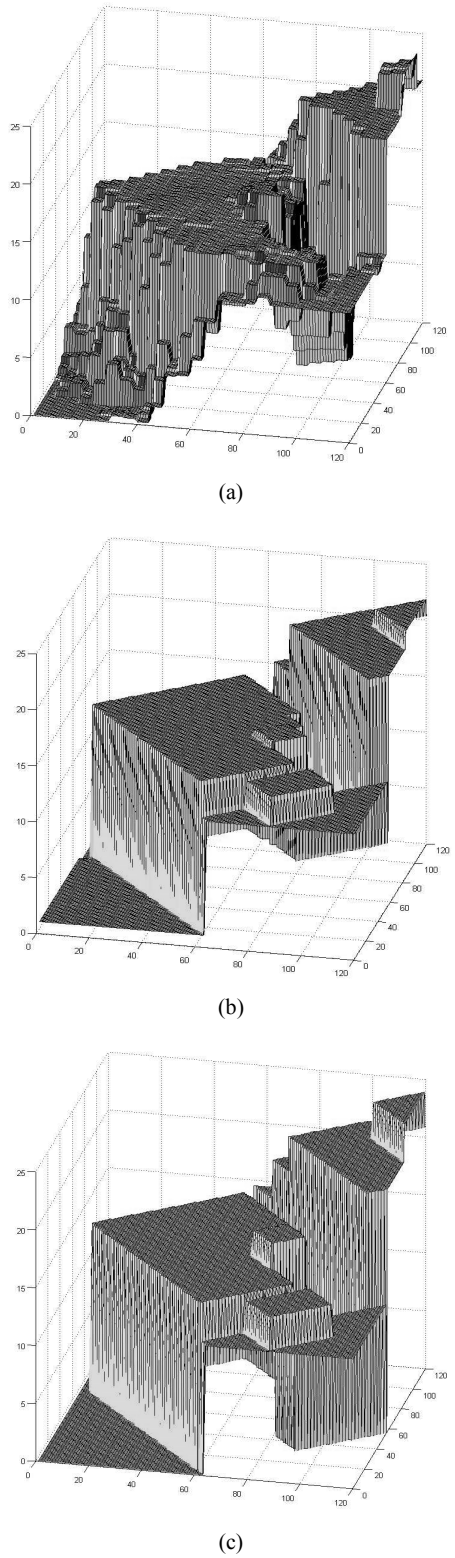
(a) (b)

Fig. 3 (a) The original elevation data (with zoom-in 5x) and (b) the upsampled elevation data

The average number of iterations for convergence of mean shift to a mode of the joint probability density was 5.57 and the maximum number of iterations was 13.

By comparing Fig. 1 (a) with Fig. 3 (b) we conclude that the well segmented optical data (prototype data for this experiment) guided very well the upsampling of the elevation data. However, although most of the segments are placed in correct elevations, there is one segment (the one that corresponds to the small square to the right of the center of Fig. 1 (a)) that has been placed at a lower level of height (close to 12m) when the true one is at 16.60m. This is due to the parameters h_s and h_e that have been selected constant throughout the whole image. Better results regarding the accuracy of the elevations are expected for the case of adaptive parameter selection, where both the h_c and h_e bandwidths are estimated from local data.

To further investigate the upsampling power of the mean shift algorithm, we first conducted experiments in order to choose the h_s and h_e parameters of both the mean shift algorithm and the best – state of the art – method, namely the bilateral upsampling. In order to quantitatively assess the accuracy of data upsampling and compare the various methods, the prototype ground truth elevation image of Fig. 5 (b) that corresponds to the original low resolution image of Fig. 5 (a) is used. As with the experiments on edge preserving smoothing of the previous section, the RMSE accuracy measure has been considered. Fig. 6 (a) shows the dependence of both algorithms on h_s and Fig. 6 (b) shows the dependence on h_e.



(a)

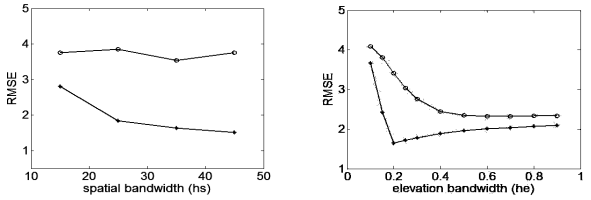
(b)

(c)

Fig. 4 The initial (a), the final (b) and the prototype (c) 3D building representations

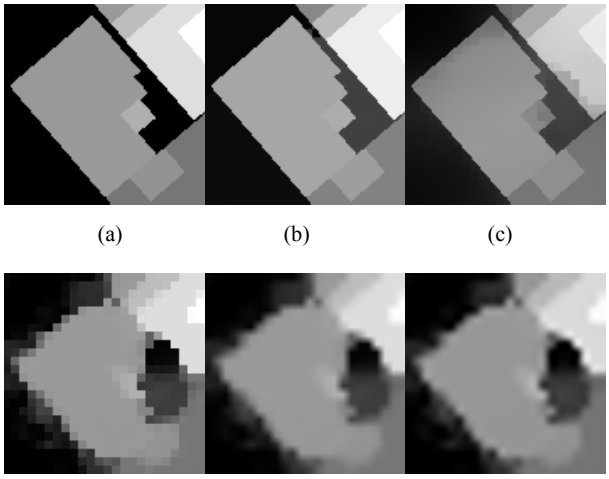
(a) (b)

Fig. 5 (a) The original elevation data (same as in Fig. 3 (a) but without the 5x zoom-in) and (b) the prototype upsampled elevation image



(a) (b)

Fig. 6 RMSE with respects to (a) spatial bandwidth h_s, and (b) elevation bandwidth h_e. The bottom lines are for mean shift and the upper lines for the bilateral algorithm



(a) (b) (c)

(d) (e) (f)

Fig. 7 (a) The prototype elevation image, (b) mean shift upsampling, (c) bilateral upsampling and, three upsampled elevation images with classic: (d) nearest neighbor, (e) bilinear, and (f) bicubic interpolation

TABLE III
COMPARISON OF FIVE UPSAMPLING INTERPOLATION METHODS

Upsampling Method	Accuracy (RMSE)
Mean Shift Algorithm	1.64
Bilateral Filtering	2.33
Nearest Neighbor Interpolation	4.38
Bilinear Interpolation	4.24
Bicubic Interpolation	4.32

Fig. 6 shows a typical instance of the various plots that can be obtained for appropriate selected values of the bandwidths. In particular, Fig. 6 (a) is obtained for $h_c = 0.1$, $h_e = 0.2$ and Fig. 6 (b) for $h_c = 0.1$, $h_s = 35$. Such plots can be used to choose good h_s and h_e parameter values. In the sequel, h_s was chosen equal to 35 for both algorithms since it gives almost the same RMSE to the choice of 45 and at the same time it preserves locality as its neighborhood includes, approximately, $70^2 = 4900$ pixels in contrast to the larger area of about 8100 pixels for the best choice. The 70 x 70 neighborhood corresponds to a 14m x 14m true square neighborhood that is more appropriate for the typical small plots of most of Athens's suberban regions. As far as h_e is concerned, the optimal values for $h_s = 35$ (see Fig. 6 (b)) were found to be 0.2 and 0.6, for the two algorithms.

The upsampled elevation data obtained with the above two algorithms as well as with the typical nearest neighbor, bilinear and bicubic interpolation methods are shown in Fig. 7.

Finally, the corresponding root mean square errors computed as deviations from the prototype elevation image are shown in Table III. It is apparent from this table that means shift gives supperior interpolation accuracy than bilateral filtering. Both the mean shift and bilateral algorithms are far more superior than other three classical interpolation methods as can also be qualitatively judged by comparing the results shown in Fig. 7.

V. 3D Building Block Reconstruction

Sections III and IV established, through comparisons, the mean shift algorithm as the most appropriate preprocessing method for 3D building reconstruction regarding both optical data smoothing and elevation data upsampling. In this Section, we employ the proposed preprocessing method for the 3D reconstruction of the whole building block.

Fig. 8 (c) shows the result of mean shift upsampling of the elevation data shown in Fig. 8 (b) for a whole building block fused with the perfectly segmented prototype image shown in Fig. 8 (a). The corresponding 3D model of the whole block is also shown in Fig. 9.

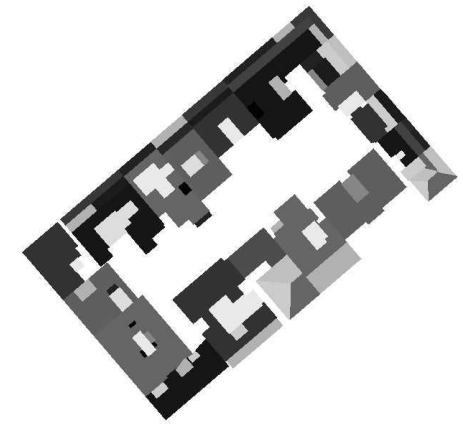
VI. Conclusion and Future Research

Comparative studies have been performed in this work to examine the edge preserving smoothing as well as the upsampling capabilities of several algorithms for the 3D building reconstruction application domain. Quantitative and qualitative results obtained using high resolution prototype segmented optical data and low resolution original elevation data show the overall superiority of the joint mean shift algorithm even against the state-of-art algorithm, namely, bilateral filtering.

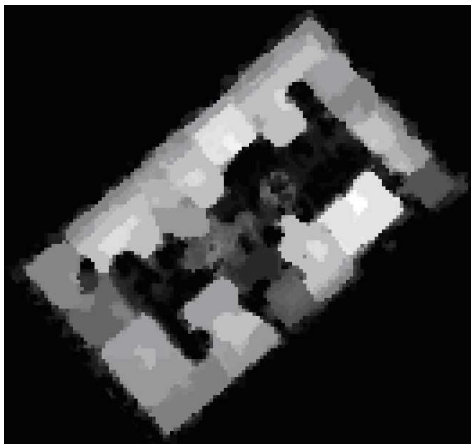
The use of the mean shift algorithm operating on the joint spatial and range (i.e. optical and elevation) domains as a preprocessing method for improving the resolution of the lidar data has been shown to constitute a promising tool towards generation of high quality 3D building models.

Further improvements on the proposed preprocessing methodology are expected by enhancing the interpolating powers of mean shift through pixelwise adaptive estimation of kernel bandwidths from local information and by

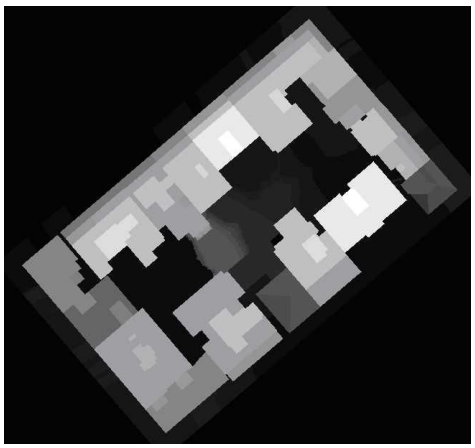
incorporating a smoothing factor on the mean shift update equation to prevent large color changes produced by long trajectories towards the nearest mode in range space.



(a)



(b)



(c)

Fig. 8 (a) The prototype segmented optical image of a building block, (b) the original elevation data after nearest neighbor upsampling, and (c) the final mean shift-based upsampled elevation data

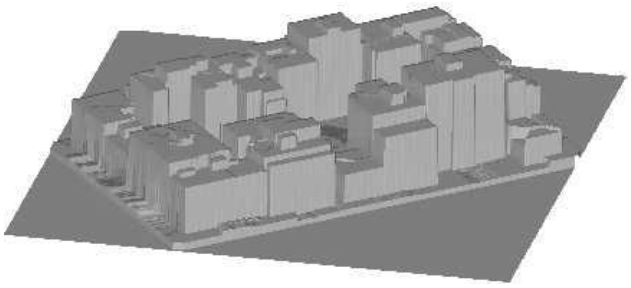


Fig. 9 The final 3D model of the building block

ACKNOWLEDGMENT

This research has been co-funded by the European Union (European Social Fund) and Greek national resources under the framework of the Archimedes III: Funding of Research Groups in T.E.I. of Athens project of the Education & Lifelong Learning Operational Programme. We would also like to thank GeoIntelligence for providing the DSM and DTM elevation data as well as the National Cadastre & Mapping Agency of Greece for providing us with the high resolution aerial photographs.

REFERENCES

[1] C. Tomasi, and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th International Conference on Computer Vision* (ICCV '98), pp. 839-846, Bombay, India, Jan. 1998.

[2] D. Comaniciu, and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 603–619, 2002.

[3] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," *Computer Graphics and Vision*, vol. 1, pp. 1-73, 2008.

[4] D. Barash, and D. Comaniciu, "A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift," *Image and Vision Computing*, vol. 22, no. 1, pp. 73-81, 2004.

[5] J. Kopf, M.F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 26, pp. 673-678, San Diego, Aug. 2007.

[6] E. Eisemann, and F. Durand, "Flash Photography Enhancement via Intrinsic Relighting, *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 23, pp. 673-678, Los Angeles, Aug. 2004.

[7] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama, "Digital Photography with Flash and No-Flash Image Pairs," *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 23, pp. 664-672, Los Angeles, Aug. 2004.

[8] J. Solomon, K. Crane, A. Butscher, and C. Wojtan, "A general framework for bilateral and mean shift filtering," CoRR, 2014.

[9] K. Fukunaga, and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on InformationTheory*, vol. 21, no. 1, pp. 32–40, 1975.

[10] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.

[11] J. van de Weijer, and R. van den Boomgaard, "Local Mode Filtering," in *Proc. Computer Vision and Pattern Recognition* (CVPR 2001), Vol. II, pp. 428-433, Hawaii, USA, Dec. 2001.