**Noname manuscript No.**
(will be inserted by the editor)

# Evaluation of sampling method effects in 3D non-rigid registration

**Marcelo Saval-Calvo · Jorge Azorin-Lopez · Andres Fuster-Guillo · Jose Garcia-Rodriguez · Sergio Orts-Escolano · Alberto Garcia-Garcia**

Received: date / Accepted: date

**Abstract** Since the beginning of 3D computer vision problems, the use of techniques to reduce the data to make it treatable preserving the important aspects of the scene has been necessary. Currently, with the new low-cost RGB-D sensors, which provide a stream of color and 3D data of approximately 30 frames per second, this is getting more relevance. Many applications make use of these sensors, and need a preprocessing to downsample the data in order to either reduce the processing time or improve the data (e.g. reducing noise or enhancing the important features). In this paper, we present a comparison of different downsampling techniques which are based on different principles. Concretely, five different downsampling methods are included: a bilinear-based method, a normal-based, a color-based, a combination of the normal and color-based samplings, and a Growing Neural Gas (GNG) based approach. For the comparison, two different models have been used acquired with the Blensor software. Moreover, to evaluate the effect of the downsampling in a real application, a 3D non-rigid registration is performed with the data sampled. From the experimentation we can conclude that depending on the purpose of the application some kernels of the sampling methods can improve drastically the results. Bilinear and GNG-based methods provide homogeneous point clouds, but color-based and normal-based provide datasets with higher density of points in areas with specific features. In the non-rigid application,

This study was supported in part by the University of Alicante and Spanish government under grants GRE11-01 and DPI2013-40534-R.

M. Saval-Calvo · J. Azorin-Lopez · A. Fuster-Guillo · J. Garcia-Rodriguez · S. Orts-Escolano · A. Garcia-Garcia
University of Alicante, San Vicente del Raspeig,03690, Spain
Tel.:(+34) 96 590 3681
Fax: (+34) 96 590 9643
E-mail: msaval@dtic.ua.es

jazorin@dtic.ua.es · fuster@dtic.ua.es · jgarcia@dtic.ua.es · sorts@dtic.ua.es · agarcia@dtic.ua.es

if a color-based sampled point cloud is used, it is possible to properly register two datasets for cases where intensity data is relevant in the model, and outperform the results if only a homogeneous sampling is used.

**Keywords** 3D downsampling · non-rigid registration · color registration

## 1 Introduction and background

The use of three dimensional data is getting more and more importance in computer vision. However, the 3D data applications commonly need a preprocessing stage to reduce the amount of data or the improve their quality. The new low-cost RGB-D sensors such as Microsoft Kinect provide depth and color information, which has made them popular in scientific and industrial environments. Nevertheless, the quality of data is not adequate for some applications. Then, it is necessary to study methods to improve the input information for the posterior processing. A common technique in many application to improve the data and reduce its dimensionality is the use of a downsampling previous step [19,16,6].

The aim of selecting a subset of points from the input could be considered from different points of view. From the point of view of reducing the amount of data, although specific architectures could be designed to improve the execution time [13], downsampling may also be seen as an improvement of the computational time, or reduction of noise. Other point of view is the maximization in the perception of the important parts (Region-of-Interest) by asymmetrically sampling the dataset. The latter could be done by sampling the points using specific features in order to have higher density of points (i.e. more samples) in the important regions for the problem. For example, normal-based sampling will reduce the number of points, keeping more data in rough and irregular parts. Nevertheless, if we are interested in reducing the amount of data to increase the processing time, a homogeneous sampling can be used.

In terms of application scope, downsampling techniques could be applied in robotics, computer vision, image processing, etc. For example, in robotics in order to construct a map of the environment in which a robot could navigate, could be necessary to reduce the points while preserving some features of the environment as planes from walls or ceilings. Other example in computer vision for object reconstruction or recognition, downsampling the 3D data extracted from a RGB-D camera, could reduce drastically the amount of noise presented in this kind of cameras.

Figures 1-A and 1-C show the selected points for an homogeneous sampling of the original image shown in Figure 1 middle-bottom and a curve of color space for a face. This face, that has been uniformly sampled, does not show any specific Region-of-Interest. Descriptive parts such as eyebrows and lips have a few quantity of points respect to the whole face (in orange). However, Figures 1-B and 1-D show the selected points for a sampling process which maximizes the number of points in descriptive regions using color information.

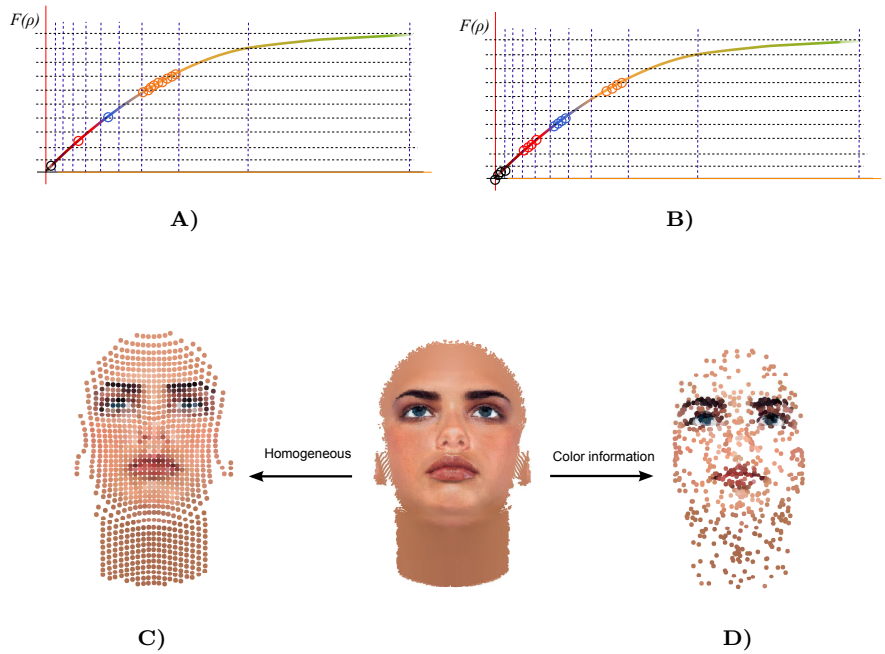


**Fig. 1** Example of downsampling effect selecting points of an image from a face. Uniform point selection represented in the curve (A) and in the image domain (C). Representative points of the face (B and D).

In 3D problems there exist different approaches for downsample the data. Mainly, we can divide them into two categories, reducing the depth map and hence reducing the final 3D points, or acting directly on the 3D point cloud. In the first group Gangwal et al. [4] used a downsampling technique in the process to provide high quality depth maps. The downsampling is applied then in the depth map not in the 3D point cloud. Concretely, they use 2D box filter with a factor, typically 4 or 8. In the paper [20] authors use a majority filter in order to sample the depth map. In [10] authors evaluated two different downsampling methods for the depth map: the original (referring to a uniform sampling) and a SVC (Scalable Video Coding [22]), concluding that SVC has a high computational complexity and based the rest of their experiments in the original method. Langmann et al. [8] studied different bilateral filtering downsampling approaches for depth images, concluding the proper rescaling results of them with the limitation of the computational time.

Regarding the second category, i.e. performing the downsample in the 3D point cloud, there exist various techniques. This category may also be divided into point-based sampling and mesh-based sampling. Based on mesh, there exist different reviews such as [2] and [9]. Recently, Castello et al. [1] presented a reduction polygonal meshes approach using view-based techniques to estimate regions where traditional methods will discard a higher number of polygons, improving the reconstruction results for visual significant surfaces. In the pa-

per [11] authors presented an efficient estimation which uses a decimation approach merging the vertices of edges which fit a condition in the process of estimating the triangles. The main disadvantage of using mesh-based sampling is that a noisy and not reliable initial point set may lead to a non-complete and incoherent mesh, which will difficult to apply any posterior method.

In the methods for 3D point-based sampling, the most common ones are the voxelization and octree-based methods [12,21,5]. However, other exist such as the work of [15] where the authors proposed a stratified point sampling which compares the uniform sampling. To evaluate the effects they use a set of test with different 3D descriptors and demonstrate the better shape descriptors provided by the point set sampled with their method. In the case of [7] the authors proposed the use of the normal angle difference with the line-of-sight and the concavity information of the point cloud to sample irregularly the point producing a shading effect, simulation the 3D human perception using shadows. Another approach was presented in [23] introducing an adaptive simplification of point cloud using neighbourhood information. Concretely, they use k-means clustering of point distribution and orientation information to scatter the data, adding boundary information and a post-processing to avoid unbalanced samplings and preserve boundary integrity. Han et al. [5] proposed a simplification of point cloud preserving edge points by using a collection of orthogonal planes passing throw each point and evaluating the number of projections to each side of the planes. If the number of projections on one plane is unbalanced regarding the sides, this point (the one which defined the plane) belongs to the edge of the model. Moreover, to reduce the points they used the orientation variation using projections (using the normal information) of each point to neighbours.

Orts-Escolano et al. presented in [18] a point cloud sampling based on artificial neural networks. Concretely, they proposed the use of Growing Neural Gas to obtain a simplification of the point cloud where each neuron of the final GNG represents one point. Moreover, they proposed a modification including the color information of each neuron using the original point cloud colors.

There exist different approaches to perform a downsampling of data, concretely this paper is focused on 3D computer vision data. Then, it is necessary to have a review that evaluates the advantages and disadvantages of the different approaches. Moreover, this review will help in future decision of which approach best fit a problem. Therefore, a comparison of different methods for 3D data downsampling is presented in this paper. In the comparison, five different techniques are going to be evaluated which include the different main categories of approaches in the state-of-the-art. The chosen methods act directly on the raw data provided by the sensor, not in a mesh, and, hence, data have not been previously altered. The data could be a point cloud or a depth image which can be later transformed into a point cloud. Moreover, although there exist different techniques, many application require a high rate of data processing, then the downsampling phase might be the fastest possible. Theses aspects motivated to choose methods which are simple in terms of computa-

tion process, and include the main two category: the depth map reduction, i.e. acting on the 2D matrix of depth information, and the 3D point set sampling.

The remaining of the paper is structured as follows: the methods are explained in Section 2. The comparison includes visual inspection in the resulting point set and the effects in a non-rigid registration problem corresponding to a possible application. The experimentation is shown in Section 3 and finally the conclusions in Section 4.

## 2 Method

In this section we introduce the different downsampling techniques that are going to be compared. To evaluate the different methods, the experimentation (Section 3) will show a possible application of the resulting data in non-rigid registration purposes. Non-rigid or deformable registration is a hot topic in 3D computer vision and, hence, its use in this comparison will provide useful conclusions for many applications. Concretely, a variant of Coherent Point Drift [14] including color in the correspondence evaluation step will be used, and briefly explained in Section 2.2. Therefore, a sampling method which stores the color information is necessary.

Five methods are going to be compared, including the different categories of point-based downsampling explained in the introductory section. In order to make more comprehensible the functioning of the methods, graphical examples are presented. The initial set of data is represented in Figure 2, where the texture or color are represented with a different shape (i.e. three colors appear including circle, pentagon and square). The orientation is represented by normal vectors.

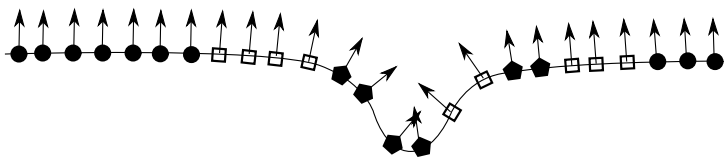

**Fig. 2** Dataset used in the example of downsampling methods. The color information is represented with different shapes (e.g. circle, square and diamond) and the orientation of the data is represented by normal vectors.

– Bilinear interpolation: a depth map reduction which interpolates the data using the linear interpolation principle, in two variables (i.e. both dimensions as depth map is a matrix of distances). Since the data is commonly provided in 2D matrices where the value of each x-y position is the depth information (Z axis), an interpolation of the 2D matrix could be performed and resulting in a smoother and smaller set of points. Figure 3 presents a generic bilinear interpolation example where a face depth map is represented in colors depending on the distance information. A region is aug-

mented in a matrix on the top-right to show a detail of the original depth values. In the bottom-right another matrix shows the sampling of the above one using bilinear interpolation, where the remaining values are the interpolated from the original.

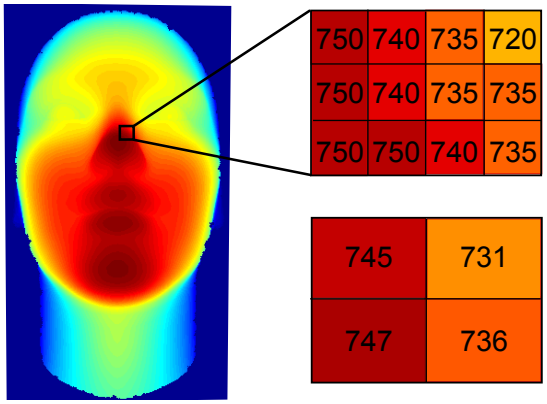


**Fig. 3** Bilinear interpolation generic example.

This method do not need color in the kernel, however the color could be interpolated as well as stored to be used in the subsequent parts (in the non-rigid registration in the case of this paper). This method is graphically explained in Figure 4 with the dataset in Figure 2, where the resulting data orientation and locations are interpolated.

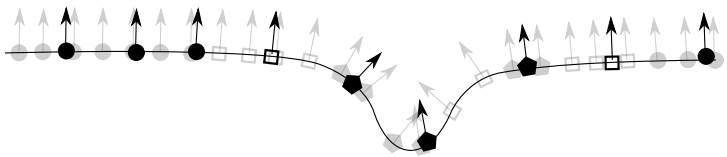

**Fig. 4** Bilinear interpolation example. The resulting data appears in bold over the gray original one.

– Normal based: this method uses the curvature of the scene to estimate the parts with higher density of points. Most sampling methods use this basis to reduce the data, mainly in applications where the mesh is going to be estimated, and then the number of triangles is reduced preserving the shape. This method uses the normals of the dataset. The process is to compare each normal with the a specific vector, defined for this case as the point of view of the camera. From this comparison we have a unidimensional variable, degrees, from 0 to 180 degrees. Since it is not possible to perceive back parts of the model from a single point of view, the maximum angle between normals and view point will be $\pm 90$. After, the space of angles

is divided into isometric segments (e.g. each 10 or 20 degree). With this division, smooth areas with small variation in the orientation will lie in the same segment, and those with large variation will occupy different segments. Finally, the segments are uniformly sampled to a predefined number of points. Therefore, as all segments are sampled by the same size, those with large number of points will keep relatively less data than those with less data, which means that smooth parts will end with less points that rough ones.

This method is graphically explained in Figure 5. In the top-left of the figure the different segments are represented by the letters corresponding to the angle parts in the circle on the bottom-left. In this case, it is clear that segments $e$ and $d$ are the ones with larger number of data. The circle has indicated the number of data in each part whose orientation lies in this angle, e.g. e - 11, f - 2. With this information we can imagine a histogram like the one represented on the right of the circle. Then, the data is uniformly sampled with 3 points or less. Hence, segments $e$ and $d$ will be downsampled to 3 and the rest, due to they do not have enough data, will remain the original. The final result on the top-right shows in bold the remaining data and in gray the original omitted data, where clearly illustrates how the area with more variation of curvature (i.e. the center) has more density of data than the smother extremes.

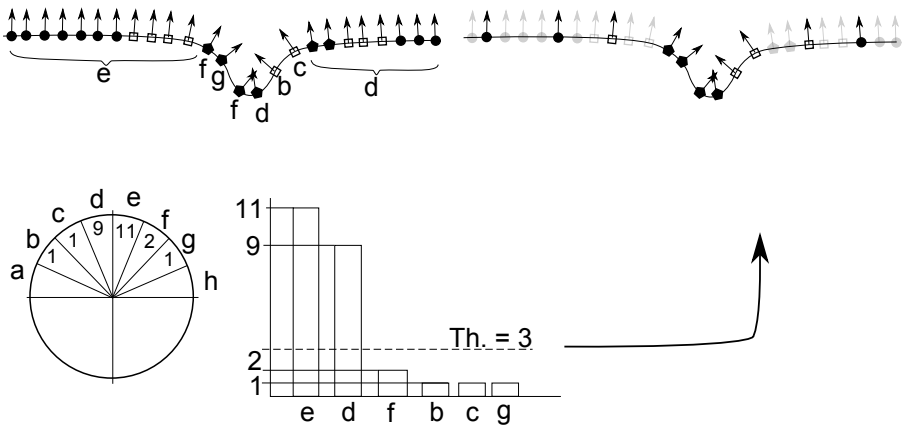


**Fig. 5** Normal downsampling. The data orientation is segmented in ranges of angles shown in the bottom-left circle, then a predefined number of data is defined (3 for this case), and eventually the result is shown on the top-right in bold over the original set in gray.

– Color based: this method is based on the same principle than the Normal based, previously explained. The color space is divided into isometric segments for each color channel. For a generic case, where the color has three channels (e.g. RGB, HSV), each of these channels are segmented and the sampling is performed separately. After, for each segment the ratio of downsampling of the channel which return more data is chosen. With this,

we avoid situations where one channel is less significant and could produce undesirable results. Once the ratio is chosen, the segments are sampled with this ratio for the three channels. An example of this is presented in the introductory section, Figure 1.
Following the example presented in the two previous methods, Figure 6 shows the result of the Color based approach. The left part shows an histogram of the amount of data per color (represented by different shapes). Once we have this information, a predefined number of points selection is set to 5 for this example. Then, each color is uniformly sampled to this value, resulting in a set of homogeneous number of data per color shown in bold, over the omitted original data in gray.

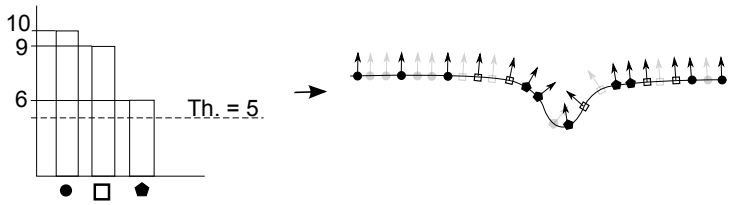


**Fig. 6** Color downsampling. The data color is segmented in ranges of colors shown in the left histogram. Then, a predefined threshold is defined (5 for this case), and eventually the result is shown on the right in bold over the original set in gray.

- Normal-color based (NC-based): the combination of both normal and color based methods returns a set of points which preserve higher density in regions of the shape with significant curvature and also those with distinctive color.
- Growing Neural Gas: the last downsampling method of the comparison is based on artificial neural networks, concretely, in the Growing Neural Gas (GNG). The idea behind this sampling is that GNG obtains a set of neurons which defines the input space. These neurons will represent points of the point cloud. Thus, if the number of neurons is lower than the actual number of points, the result is a downsample of the three dimensional space. The GNG based sampling technique is explained in detail in the subsection 2.1. Moreover, a variant which estimates the color of the neurons has been proposed and used in [17]. This method is graphically explained in Figure 7. The figure shows on the left side an original set of data, concretely, a set of points that describes a surface. The data is then sampled by neurons (ellipses) in the right image and the connexions between neurons.

## 2.1 Sampling GNG-based

A brief explanation of the GNG neural network is presented. Despite the use in this work is focused on 3D point sampling, we are going to present the general steps included in the process of learning.

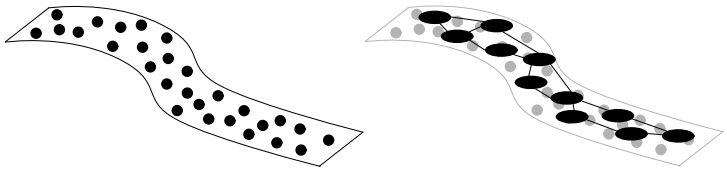

**Fig. 7** Growing neural gas downsampling.

The Growing Neural Gas (GNG) [3] is an artificial neural network which is characterized by a growth process that takes place from minimal network size, and new neurons are inserted successively using a particular type of vector quantization. In order to evaluate the location of a new neuron, a local error is estimated in the adaptation process, and inserting the new neuron in the region of the highest error. The connection between neurons is carried out using the Competitive Hebbian Learning (CHL) algorithm. The process finishes when a condition is fulfilled, e.g. optimal topology is achieved or number of neurons is reached.

The network is specified as:

- A set $N$ of nodes (neurons). Each neuron $c \in N$ has its associated reference vector $w_c \in R^d$. The reference vectors can be regarded as positions in the input space of their corresponding neurons.
- A set of edges (connections) between pairs of neurons. These connections are not weighted and its purpose is to define the topological structure. An edge aging scheme is used to remove connections that are invalid due to the motion of the neuron during the adaptation process.

The GNG learning algorithm is as follows:

1. Start with two neurons $a$ and $b$ at random positions $w_a$ and $w_b$ in $R^d$.
2. Generate at random an input pattern $\xi$ according to the data distribution $P(\xi)$ of each input pattern.
3. Find the nearest neuron (winner neuron) $s_1$ and the second nearest $s_2$.
4. Increase the age of all the edges emanating from $s_1$.
5. Add the squared distance between the input signal and the winner neuron to a counter error of $s_1$ such as:

$$\triangle error(s_1) = \|w_{s_1} - \xi\|^2 \tag{1}$$

6. Move the winner neuron $s_1$ and its topological neighbors (neurons connected to $s_1$) towards $\xi$ by a learning step $\epsilon_w$ and $\epsilon_n$, respectively, of the total distance:

$$\triangle w_{s_1} = \epsilon_w(\xi - w_{s_1}) \tag{2}$$

$$\triangle w_{s_n} = \epsilon_n(\xi - w_{s_n}) \tag{3}$$

   For all direct neighbors $n$ of $s_1$.
7. If $s_1$ and $s_2$ are connected by an edge, set the age of this edge to 0. If it does not exist, create it.

8. Remove the edges larger than $a_{max}$ . If this results in isolated neurons (without emanating edges), remove them as well.
9. Every certain number $\lambda$ of input patterns generated, insert a new neuron as follows:
   - Determine the neuron $q$ with the maximum accumulated error.
   - Insert a new neuron $r$ between $q$ and its further neighbor $f$:

$$w_r = 0.5(w_q + w_f) \tag{4}$$

   - Insert new edges connecting the neuron $r$ with neurons $q$ and $f$, removing the old edge between $q$ and $f$.
10. Decrease the error variables of neurons $q$ and $f$ multiplying them with a consistent $\alpha$. Initialize the error variable of $r$ with the new value of the error variable of $q$ and $f$.
11. Decrease all error variables by multiplying them with a constant $\gamma$.
12. If the stopping criterion is not yet achieved (in our case the stopping criterion is the number of neurons), go to step 2.

The GNG-based sampling method also includes other benefits, such as the capability to filter noise in the 3D point cloud preserving the topology of the input space in the output representation [18,17].

A further improvement has been done by including the color information of the original point cloud in the neurons for the representation after the sampling process. After the process of learning each neuron takes the color from the nearest neighbours, which results in an interpolated value of the surrounding points.

### 2.2 Non-rigid registration using color

In order to evaluate the effects of the sampling methods, a non-rigid registration application is going to be used, topic which is currently under a intense study and will allow to extract useful conclusions. For the sake of completeness, a brief description of the registration method is presented in this section.

In order to evaluate the effects of the sampling with non-rigid registration of 3D point clouds, a variant of CPD [14] which takes into account the color information is used, hereafter referred as Color-CPD. Original CPD uses point set space position to register both point clouds. Using Gaussian Mixture Models with Expectation-Maximization, the algorithm iteratively aligns them.

Given two point sets where $Y_{MxD} = (y1, \cdots, y_M)$ is the moving set which will be aligned with the static point set $X_{NxD} = (x1, \cdots, x_N)$, where D is the dimensionality. Each point in $Y$ is the mean of a Gaussian Mixture Model (GMM), all with equal isotropic covariance $\sigma^2$. The objective is to find the parameters of the GMM, mainly the mean as it will define the position of points, making $Y$ fit $X$. This process will be carried out by maximizing the likelihood, or equivalently, minimizing the negative log-likelihood. In order to do this, Expectation-Maximization is used, using Bayes' theorem to find the current, or old, parameters by means of the posterior probability.

CPD uses for the posterior probability ($P$) the space position of the points. In this, the Color-CPD, we introduced the intensity information, combining it with the space position to calculate this probability. To do this, we use another GMM for the color instead of including the color in X and Y, making them 6D (3D + RGB). The union of both probabilities (PL and PC, location and color respectively) are performed and weighted to let the algorithm adapt different situations. The final posterior probability is:

$$P_{(m|x_n)} = \frac{(PL_{m|x_n})^{wl} * (PC_{m|x_n})^{wc}}{(\sum_{k=1}^{M} PL_{k|x_n})^{wl} * (\sum_{j=1}^{M} PC_{j|x_n})^{wc} + o_L + o_C}$$

$wl$ and $wc$, $wl+wc = 1$ are the weights to balance the effect of each probability. $o_L$ and $o_C$ are other gaussian probabilities to take into account outliers in 3D position and color intensity.

$$PL_{m|x_n} = exp^{-\frac{(\|x_n - y_m\|^2)}{(2\sigma^2)}}$$

$$PC_{m|x_n} = exp^{-\frac{(\|cx_n - cy_m\|^2)}{(2\sigma_c^2)}}$$

$\sigma_c^2$ represents the covariance of the color space, different from $\sigma^2$.

Lastly, it is very important to highlight that the method uses Coherent Motion Theory to regularize the movement of the points, avoiding non-sense displacements.

## 3 Experimental Results

In this section it is presented the evaluation of the different downsampling methods. On the one hand evaluating the resulting point sets to observe the effects of each technique (Subsection 3.1). On the other hand an evaluation with the previously explained non-rigid registration method will take place to analyse the resulting sampled point sets in a real application (Subsection 3.2).

The dataset includes two different objects: a flower and a face. The synthetic models have been acquired using the Blensor tool, a Blender plugin which simulates a Microsoft Kinect RGB-D sensor. The models have been deformed with Blender to have three shapes, origin, small deformation and large deformation. The comparative of sampling methods includes the five different sampling techniques explained before (Section 2).

Figure 8 and 9 show the face and flower models used for the experiments. The images are from left to right: the target, a first deformation, and a second larger deformation. The face deformations could be seen as elastic deformations, because the face shape remains the same global size but certain parts suffer some displacements. Concretely, the first deformation is a eyebrow rise and a mouth change. The second moves both eyebrows and the mouth, changes the nose and the chin. For the flower, it could be seen as growth deformations due to the size of the object changes. The first deformation enlarges a little the leaves and the second is a larger deformation.

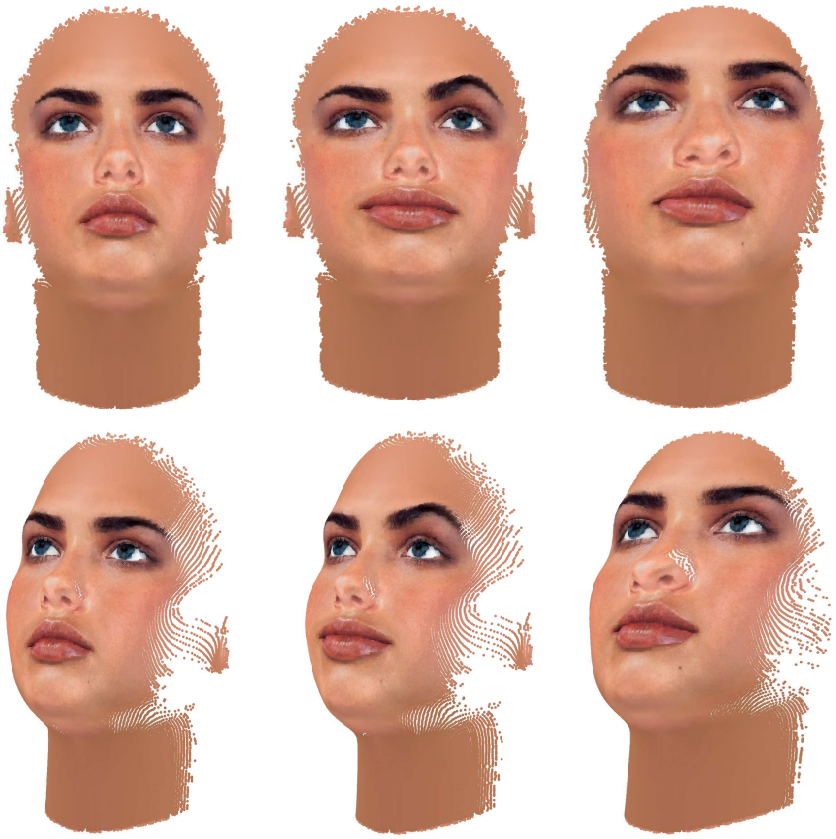

**Fig. 8** The face model used in the experimentation. Two viewpoints (each per row) of the faces used. From left to right, the original face shape as target for the deformations in second and third columns.

### 3.1 Evaluation of downsampling methods

Using the two models (i.e. face and flower), the five different techniques of data sampling have been evaluated, using three sampling rates (250, 500 and 1000 points). Different results are expected depending on the feature to be considered of the sampling method.

Figure 10 and Figure 11 present the result of the different downsampling techniques for 1000 points. In the figures, it is shown the target model (first row), first deformation (middle row), and the larger second deformation (bottom row). From left to right, the sampling techniques presented are: bilinear, normal-based, color-based, NC-based, and GNG. The same experimentation has been carried out for the other two sampling rates, however only this one is presented as an example of the results provided by the different techniques.

From the data presented in Figures 10 and 11, bilinear and GNG downsampling techniques provide the expected spatial data distribution, as both

**Fig. 9** The flower model used in the experimentation. Two viewpoints (each per row) of the flower used. From left to right, the original flower shape as target for the deformations in second and third columns.

are conceived for sampling the data homogeneously along the space, as it is possible to visually appreciate for both models in the figures. This procedure is interesting when the purpose of the method that uses these data needs to have information about the whole space, such as rendering or surface reconstruction. Moreover, GNG has demonstrated noise reduction capabilities in [18]. The normal-based sampling returns higher density in those parts with higher shape variation. The face clearly shows higher density in the nose sides and eyes cavity. For the flower, parts with orientation different from the rest has higher density of data. This reveals that the reference vector to extract the angle is a critical aspect which must be taken into account. In the second deformation of the flower, the top part has higher density because it look frontally and only a yellow part of the rest looks to the same orientation. Similar situation occurs with the right part. The color-based sampling for the face returns more data in eyebrows, eyes and nose. For the flower, the color-based sampling data has a higher density in the center region due to it is the most different one, similar to the stem. The combination NC-based provides a hybrid solution of both previous methods.

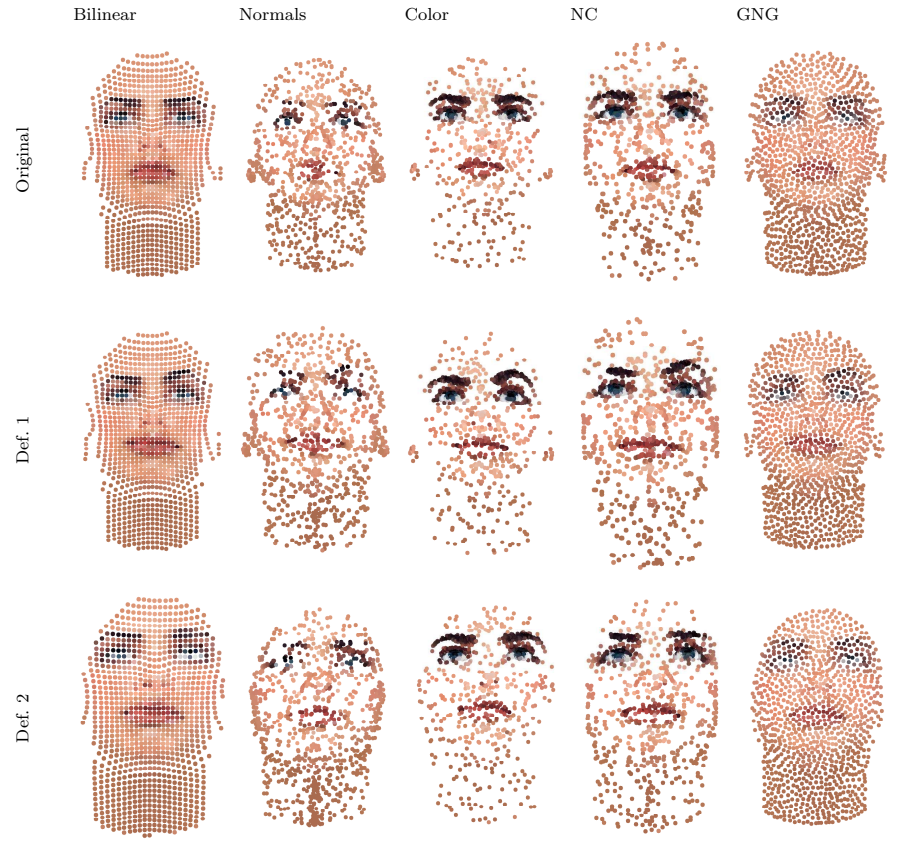


**Fig. 10** Sampling examples for 1000 data points. First row shows the original,second and third the two deformation (small and large) of face. From left to fight the sampling techniques are bilinear, normal, color, NC-based, and GNG.

As general comparative, color-based, normal-based and NC-based provide very detailed data in those parts with a specific characteristics. However, the main drawback is that in cases where the whole shape wants to be used, they do not provide proper data. This is the opposite in bilinear and GNG based methods which return homogeneous data spread on the surface of the subject.

### 3.2 Non-rigid registration evaluation

In order to evaluate the effect of the output data from the downsampling method, a non-rigid registration experimentation is presented in this section. As the color is a key feature in many 3D applications, this evaluation is going to be carried out using a non-rigid registration with only 3D information (Coherent Point Drift [14]) and a variant including the color in the registration process.

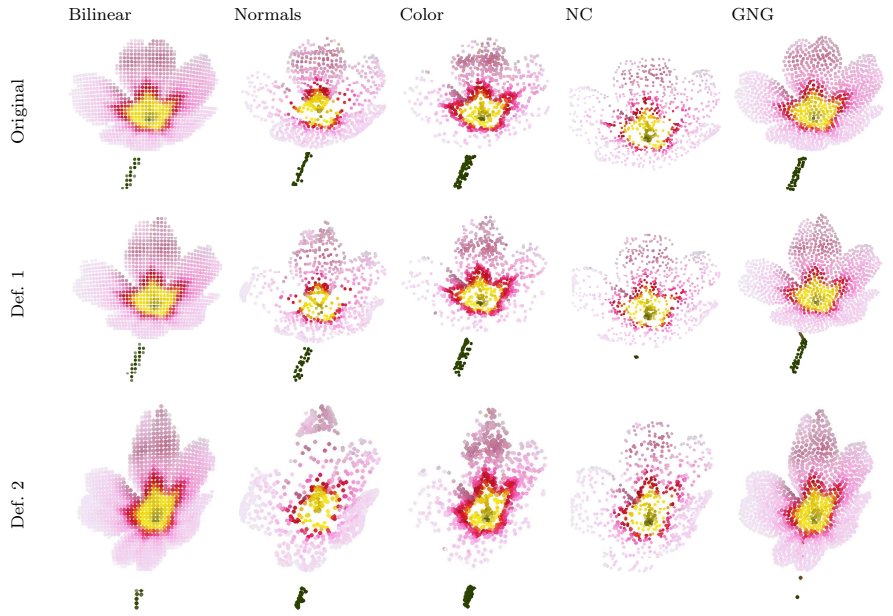


**Fig. 11** Sampling examples for 1000 data points. First row shows the original, second and third the two deformations (small and large) of the flower. From left to fight the sampling techniques are bilinear, normal, color, NC-based, and GNG.

A comparative evaluation of Color-CPD and CPD is here presented for registering synthetic realistic subjects. The color information, used by Color-CPD, allows the registration method to achieve good results in accuracy when the surface is not very detailed so the drift of points is not constrained by the irregularities of the shape.

Using the data sampled in the Subsection 3.1, the non-rigid registration methods are qualitatively evaluated by visual inspection. Similarly to the previous section, here the 1000 point sampling ratio is depicted as an example, in spite of the 250 and 500 has been also performed to corroborate the results. Figure 13 shows the face shape for Color-CPD and the original CPD with 1000 points respectively. Moreover, Figure 15 shows the flower shape for Color-CPD and the original CPD with the same point sampling. Concretely, figures show the registration for the second deformation of each shape as it is the larger one. For each figure, the first row presents the Color-CPD method and the second the original CPD. From left to right, the sampling techniques are: bilinear, normal-based, color-based, NC-based, and GNG.

To later analyse the registration, we pay special attention to a specific Region-of-Interest (ROI) for each model (i.e. those parts that are the aim of the study), shown in Figure 12. In the face, the ROI will correspond to the mouth and eyebrows due to they are the parts which are mainly displaced. The ROI in the flower will correspond to the central part, pink and yellow, as they do not deform in color accordingly to the rest of the leaves (i.e. the deformation

produces an enlargement of the tip of leaves, but the center remains the same). This simulates the growth of a flower, where not all parts grow in the same way.

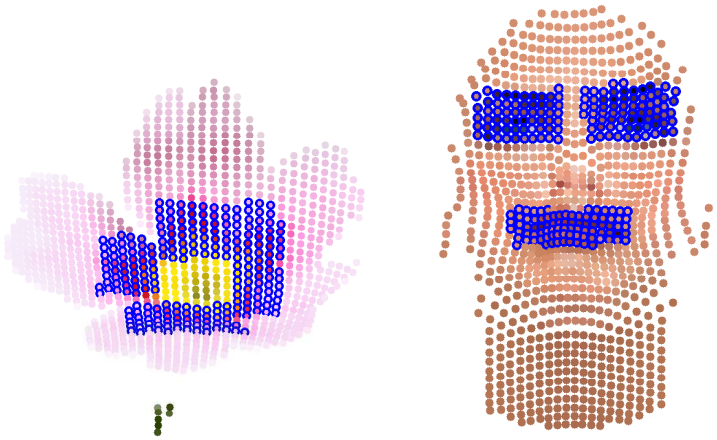

**Fig. 12** Example of Region of Interest for both shapes. The ROI are highlighted with blue circles.

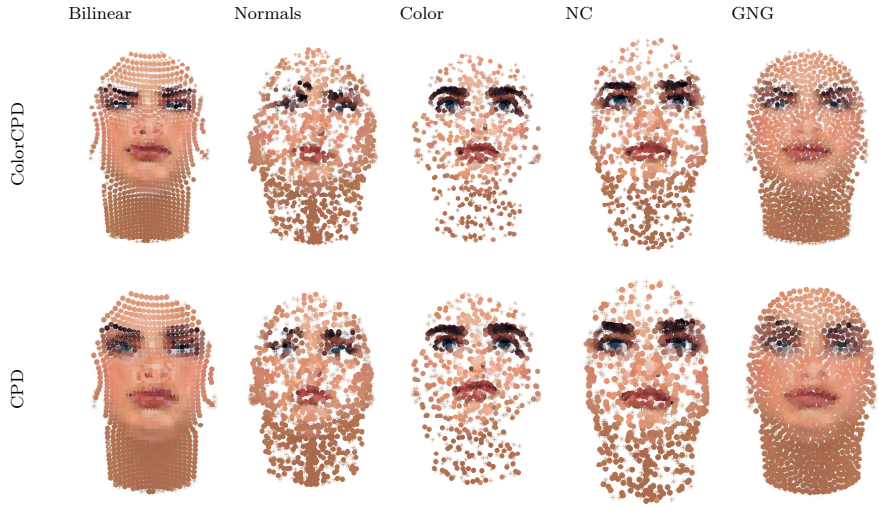


**Fig. 13** Non-rigid registration result of face shape for a 1000 points sampling. The first row shows color CPD, and the original CPD in the second. Columns are from left to right, bilinear, normal-based, color-based, NC-based, GNG.

We can conclude different aspects from the results of the experimentation. The most important one is that the sampled data provided to the registration

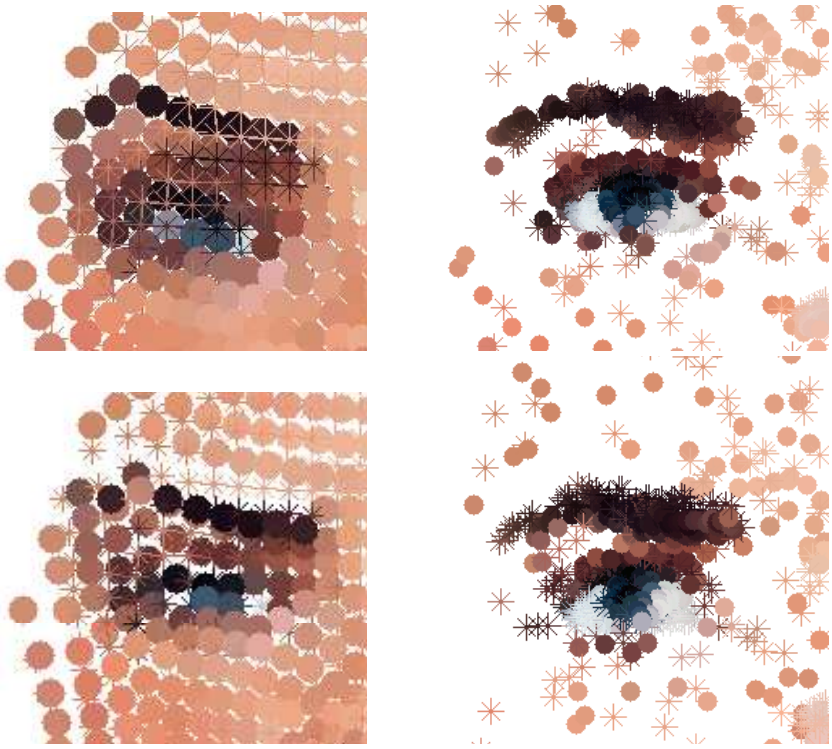

**Fig. 14** Enlarged example of the ROI for the face sampled with 1000 points with bilinear (left) and color-based (right). The first row shows the original CPD and the second the Color-CPD variant.

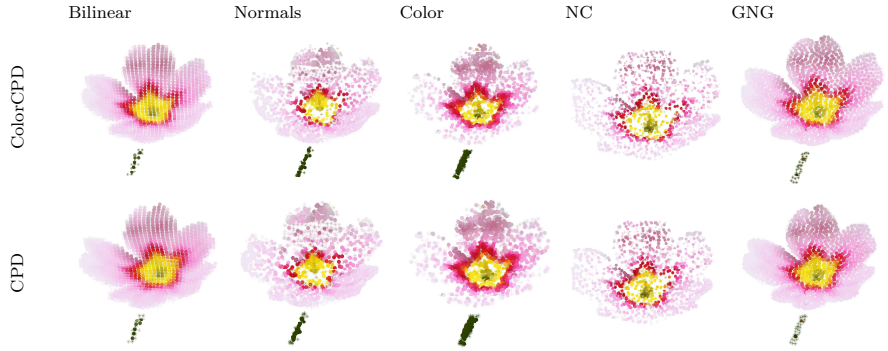


**Fig. 15** Non-rigid registration result of flower shape for a 1000 points sampling. The first row shows color CPD, and the original CPD in the second. Columns are from left to right, bilinear, normal-based, color-based, NC-based, GNG.

method has a direct effect in the final alignment. This occurs notoriously in the face because the ROI (eyebrows and mouth) performs the main movement of the deformation. Color-CPD aligns the points properly (i.e. they align the

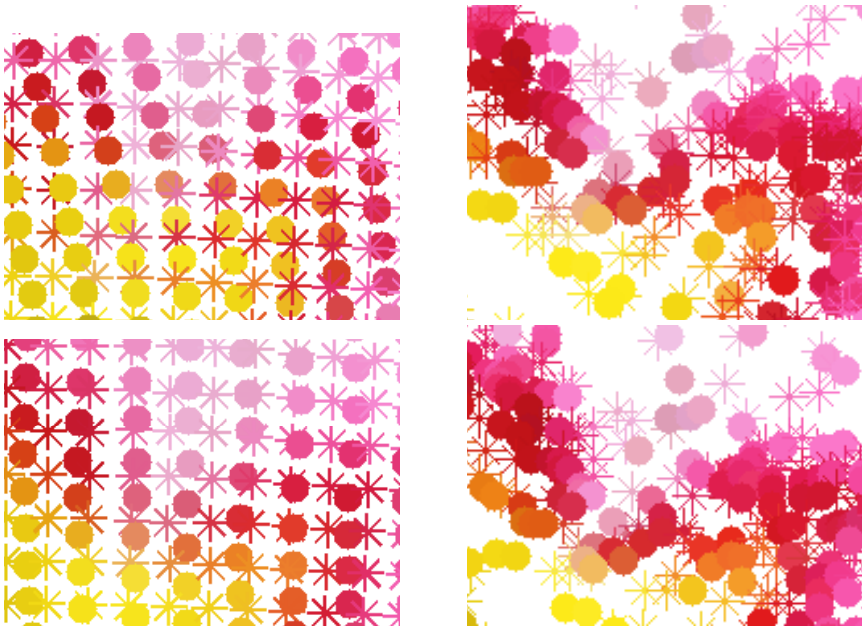

**Fig. 16** Enlarged example of the ROI for the flower sampled with 1000 points with bilinear (left) and color-based (right). The first row shows the original CPD and the second the Color-CPD variant.

point in the expected way by color and location) while CPD misaligns as the registered points align to points with different colors. However, using color-based and NC-based, both Color-CPD and CPD properly align due to the distribution of points has higher density in the ROI (higher density of points in ROI). Then, if mainly points in the ROI are used, being this part the one that moves, original CPD is indirectly constraint by the color, so both algorithms will perform similar results. Regarding the rest of sampling methods, Color-CPD achieves better alignment of the points in color and location space.

Figure 14 shows the ROI of the face for bilinear and color-based to clearly visualize that the registration quality improves with the data provided by the second method.

The flower shape presents a similarly behaviour to the face in the registration results. When the data comes from either color-based or NC-based, both Color-CPD and CPD achieve similar results. Moreover, when the data has been sampled using GNG or bilinear, the proposed Color-CPD achieves higher registration accuracy than the original CPD. As the deformation in this shape is not isometric, being the tips of some leaves the parts that get larger compared to other. Color-CPD moves the points differently in the tip of the leaves than the ROI, achieving accurate results. However, as CPD moves coherently, the points shrink all together (the registration is from the larger to the original position) producing that ROI ends in a wrong color alignment.

Figures 16 shows a detail of the registration to evaluate visually the accuracy of both methods with data sampled with bilinear and color-based. It is easy to appreciate that both Color-CPD and original CPD achieve proper results with the sampling that provide higher density in the ROI.

The evaluation presented has been done for three different sampling ratios. The one shown before corresponds to the sampling which provides 1000 data, however, the same behaviour has been observed for the other two ratios. Actually, in the Section 3.3, Table 1 shows the registration for the different cases. For all cases, the use of a downsampling method which internally uses a specific feature helps in posterior processes, being the color in the case of the registration performed.

### 3.3 Quantitative results

In order to provide a quantitative analysis of the effects of downsampling in the registration process, this section presents a summary of the registration for the whole set of the performed experiments. It includes the two deformations of the plant and the face, with the three levels of sampling and the five evaluated methods. In order to provide reliable results, the error has been obtained using just the points belonging to the ROI (eyes and mouth in the face shown in Figure 12-right and center of the plant in Figure 12-left). The results are presented in Table 1 where the first column represents the downsampling method, the second the sampling points, and next all error per registration method, shape, and deformation.

In order to provide a robust measure, the error takes into account the color information. The quality of the registration is related to the real correspondence. Correspondences just in location space are not representative by themselves. However, they become useful information if we observe the correspondence in the color space, i.e. the closest point in the location space is not a good correspondence if the colors are different. In the same way, this measure avoids the problem that happens when one point may be registered in between various points as not the same distribution of data in the location and color is obtained after the sampling, e.g. in the flower one deformation is enlarging, then for the same number of points sampled one leaf will have more spread points than the smaller.

Therefore, first, we find the set of closest points $P_j$ (Eq. 5) as the points $p_i$ in $X$ for a point $y_j$ within a defined distance $r_\epsilon$ in the location space ($\|\cdot\|_L$ represents the euclidean distance):

$$P_j = \{p_i, \|p_i - y_j\|_L \leq r_\epsilon \quad | \quad \forall i \in X\} \tag{5}$$

The distance $r_\epsilon$ (Eq. 6) is calculated as two times the average of the minimum distances $D$ (Eq. 7) between each point in $Y$ to the dataset $X$:

$$r_\epsilon = 2 \cdot \frac{\sum_{i \in D} d_i}{n_X} \tag{6}$$

$$D = \{d_i = min(\{\|x_i - y_j\|_L, \forall j \in Y\}) \quad | \quad \forall i \in X\} \tag{7}$$

, being $n_X$ the elements in $X$.

After, the error $\epsilon_j$ (Eq. 8) is the difference in the color space between the point $y_j$ and the one of the closest $p_i$ which minimizes the distance in color:

$$\epsilon_j = min\{\|p_i - y_j\|_C\} \forall i \in P_j \tag{8}$$

Finally, the error $\epsilon$ (Eq. 9) is the mean of all $\epsilon_j$, being $n_Y$ the elements in $Y$.

$$\epsilon = \frac{\sum_{\forall j \in Y} \epsilon_j}{n_Y} \tag{9}$$

Taking into account the $\mu_D$ as the mean value of the error $\epsilon$ per deformation, we can conclude that Color-CPD works better in general terms than the original version.In average, the error for the Color-CPD is 6.07, while the original CPD achieves 8.30 for the different experiments. However, since the aim of this paper is to evaluate the downsampling methods, we are going to focus on the average error for the different downsampling methods according to the Color-CPD ($\mu_C$) and the original CPD ($\mu_O$). In both cases, the color-based downsampling method achieves the lowest error, being the original CPD where it is more notorious as the method do not use it in the registration process.

**Table 1** Table of registration results for the color information for all experiments. The $\mu_D$ bottom row represents the mean of the values per deformation, the $\mu_C$ and $\mu_O$ are the means for each downsampling method for Color-CPD and Original CPD respectively.

| | | Color-CPD | | | | | CPD | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Face | | flower | | | Face | | flower | | |
| Method | Sampling | Def1 | Def2 | Def1 | Def2 | $\mu_C$ | Def1 | Def2 | Def1 | Def2 | $\mu_O$ |
| Bilinear | 1000 | 3.50 | 5.38 | 4.54 | 4.42 | 4.46 | 3.97 | 17.49 | 11.25 | 23.00 | 13.93 |
| Color | 1000 | 2.71 | 2.99 | 6.61 | 3.98 | **4.07** | 2.07 | 2.31 | 5.42 | 3.36 | **3.29** |
| GNG | 1000 | 4.48 | 3.62 | 5.09 | 5.21 | 4.60 | 6.72 | 8.73 | 7.28 | 16.30 | 9.76 |
| NC | 1000 | 3.74 | 5.36 | 7.93 | 5.97 | 5.75 | 2.58 | 3.73 | 7.83 | 8.12 | 5.56 |
| Normal | 1000 | 4.03 | 5.61 | 5.88 | 12.23 | 6.94 | 6.71 | 4.09 | 3.80 | 8.08 | 5.67 |
| Bilinear | 500 | 0.42 | 3.27 | 9.94 | 7.18 | 5.20 | 2.25 | 8.27 | 11.96 | 19.28 | 10.44 |
| Color | 500 | 1.97 | 4.12 | 4.38 | 4.00 | **3.62** | 2.29 | 2.53 | 4.15 | 3.37 | **3.08** |
| GNG | 500 | 4.96 | 4.49 | 5.32 | 5.98 | 5.19 | 5.25 | 8.20 | 9.72 | 15.78 | 9.74 |
| NC | 500 | 3.50 | 5.13 | 6.10 | 15.85 | 7.65 | 2.36 | 3.23 | 5.59 | 13.07 | 6.06 |
| Normal | 500 | 6.73 | 7.36 | 6.01 | 7.23 | 6.84 | 11.13 | 7.92 | 5.40 | 10.25 | 8.67 |
| Bilinear | 250 | 3.98 | 12.03 | 7.12 | 10.70 | 8.46 | 4.00 | 18.28 | 22.20 | 14.99 | 14.87 |
| Color | 250 | 4.12 | 7.40 | 7.41 | 5.56 | **6.12** | 2.89 | 2.57 | 5.15 | 10.63 | **5.31** |
| GNG | 250 | 7.06 | 5.39 | 7.39 | 8.32 | 7.04 | 8.41 | 7.53 | 7.38 | 18.23 | 10.39 |
| NC | 250 | 7.13 | 7.66 | 7.99 | 9.37 | 8.04 | 3.32 | 4.36 | 7.81 | 12.86 | 7.09 |
| Normal | 250 | 5.06 | 9.05 | 5.49 | 8.82 | 7.11 | 8.77 | 12.17 | 9.28 | 12.25 | 10.62 |
| $\mu_D$ | | **4.23** | **5.92** | **6.48** | **7.65** | **6.07** | 4.85 | 7.43 | 8.28 | 12.64 | 8.30 |

## 4 Conclusion

A comparison of five different downsampling methods for 3D data has been presented. For downsample 3D data there exist different approaches, including 2D depth map sampling, point-based, and mesh-based. The mesh is obtained triangulating the surface from the points, hence we have focused this paper on the two first categories due to they use the data provided directly from the sensor. The comparison has been carried out using a bilinear method for depth map reduction, a normal-based technique which uses the orientation information to estimate the sampling ration. Moreover, a color-based has been evaluated which provides a sampling where homogeneous color regions are highly reduced and maintains higher density of point in descriptive parts. A combination of normal and color-based techniques has been also included in the comparison. Finally, a Growing Neural Gas sampling method has been used. The experimentation used two models, a face and a flower. For the visual inspection, the GNG and the bilinear provide homogeneous sampling which allow to have information of the whole model. However, normal, color and the combination provide different densities of points depending on the information used in the kernel of method. This could help in the posterior processing of the data, since it has been demonstrated in a second experiment where a non-rigid registration has been done. The registration experiments where focused on aligning the models with different deformations, where the alignment quality is evaluated with both color and point location results. For this case, the color-based sampling models achieve the best registration. For future works, we are evaluating more methods based on different features and also we will include mesh-based methods in the comparison.

## References

1. Castelló, P., Chover, M., Sbert, M., Feixas, M.: Reducing complexity in polygonal meshes with view-based saliency. Computer Aided Geometric Design **31**, 279–293 (2014). DOI 10.1016/j.cagd.2014.05.001
2. Cignoni, P., Montani, C., Scopigno, R.: A comparison of mesh simplification algorithms. Computers & Graphics **22**, 37–54 (1997)
3. Fritzke, B.: A Growing Neural Gas Network Learns Topologies, vol. 7, pp. 625–632. MIT Press (1995)
4. Gangwal, O.P., Berretty, R.P.: Depth map post-processing for 3D-TV. In: Digest of Technical Papers - IEEE International Conference on Consumer Electronics (2009). DOI 10.1109/ICCE.2009.5012253
5. Han, H., Han, X., Sun, F., Huang, C.: Point cloud simplification with preserved edge based on normal vector. Optik - International Journal for Light and Electron Optics (2015). DOI 10.1016/j.ijleo.2015.05.092. URL http://linkinghub.elsevier.com/retrieve/pii/S0030402615004052
6. Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D.: RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments (2012). DOI 10.1177/0278364911434148
7. Katz, S., Tal, A.: Improving the visual comprehension of point sets. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 121–128 (2013). DOI 10.1109/CVPR.2013.23

8. Langmann, B., Hartmann, K., Loffeld, O.: Comparison of depth super-resolution methods for 2d/3d images. International Journal of Computer Information Systems and Industrial Management Applications **3**, 635–645 (2011)
9. Lee, K.H., Woo, H., Suk, T.: Point Data Reduction Using 3D Grids (2001). DOI 10.1007/s001700170075
10. Li, Y., Sun, L.: A novel upsampling scheme for depth map compression in 3DTV system. In: 28th Picture Coding Symposium, PCS 2010, pp. 186–189 (2010). DOI 10.1109/PCS.2010.5702456
11. Mamou, K., Ghorbel, F.: A simple and efficient approach for 3D mesh approximate convex decomposition. Image Processing (ICIP), 2009 16th IEEE International Conference on (2009). DOI 10.1109/ICIP.2009.5414068
12. McNeely, W.A., McNeely, W.A., Puterbaugh, K.D., Puterbaugh, K.D., Troy, J.J., Troy, J.J.: Six degree-of-freedom haptic rendering using voxel samping. Comput. Graph. (SIGGRAPH Proc.) pp. 401–408 (1999). DOI 10.1145/1198555.1198605
13. Mora-Mora, H., Mora-Pascual, J., García-Chamizo, J.M., Jimeno-Morenilla, A.: Real-time arithmetic unit. Real-Time Systems **34**(1), 53–79 (2006)
14. Myronenko, A., Song, X.: Point set registration: coherent point drift. IEEE transactions on pattern analysis and machine intelligence **32**(12), 2262–75 (2010). DOI 10.1109/TPAMI.2010.46
15. Nehab, D., Shilane, P.: Stratified point sampling of 3d models. In: Proceedings of the First Eurographics Conference on Point-Based Graphics, SPBG'04, pp. 49–56. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2004). DOI 10.2312/SPBG/SPBG04/049-056. URL http://dx.doi.org/10.2312/SPBG/SPBG04/049-056
16. Newcombe, R.A., Davison, A.J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., Molyneaux, D., Hodges, S., Kim, D., Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality, pp. 127–136. IEEE (2011). DOI 10.1109/ISMAR.2011.6092378
17. Orts-Escolano, S., Garcia-Rodriguez, J., Morell, V., Cazorla, M., Garcia-Chamizo, J.: 3d colour object reconstruction based on growing neural gas. In: Neural Networks (IJCNN), 2014 International Joint Conference on, pp. 1474–1481 (2014)
18. Orts-Escolano, S., Morell, V., Garcia-Rodriguez, J., Cazorla, M.: Point cloud data filtering and downsampling using growing neural gas. In: The 2013 International Joint Conference on Neural Networks, IJCNN 2013, Dallas, TX, USA, August 4-9, 2013, pp. 1–8 (2013)
19. Ryde, J., Hu, H.: 3D mapping with multi-resolution occupied voxel lists. Autonomous Robots **28**, 169–185 (2010). DOI 10.1007/s10514-009-9158-3
20. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. **1** (2003). DOI 10.1109/CVPR.2003.1211354
21. Schnabel, R., Klein, R.: Octree-based Point-Cloud Compression. In: SPBG, pp. 111–120 (2006). DOI 10.2312/SPBG/SPBG06/111-120
22. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the scalable video coding extension of the H.264/AVC standard. In: IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, pp. 1103–1120 (2007). DOI 10.1109/TCSVT.2007.905532
23. Shi, B.Q., Liang, J., Liu, Q.: Adaptive simplification of point cloud using k-means clustering. CAD Computer Aided Design **43**, 910–922 (2011). DOI 10.1016/j.cad.2011.04.001