

Article

## Assessment and Calibration of a RGB-D Camera (Kinect v2 Sensor) Towards a Potential Use for Close-Range 3D Modeling

Elise Lachat \*, Hélène Macher, Tania Landes and Pierre Grussenmeyer

ICube Laboratory UMR 7357, Photogrammetry and Geomatics Group, National Institute of Applied Sciences (INSA Strasbourg), 24, Boulevard de la Victoire, 67084 Strasbourg Cedex, France;  
E-Mails: helene.macher@insa-strasbourg.fr (H.M.); tania.landes@insa-strasbourg.fr (T.L.);  
pierre.grussenmeyer@insa-strasbourg.fr (P.G.)

\* Author to whom correspondence should be addressed; E-Mail: elise.lachat@insa-strasbourg.fr;  
Tel.: +33-388-14-47-33; Fax: +33-388-14-47-33.

Academic Editors: Diego Gonzalez-Aguilera, Devrim Akca, Richard Müller and Prasad S. Thenkabail

Received: 30 June 2015 / Accepted: 23 September 2015 / Published: 1 October 2015

---

**Abstract:** In the last decade, RGB-D cameras - also called range imaging cameras - have known a permanent evolution. Because of their limited cost and their ability to measure distances at a high frame rate, such sensors are especially appreciated for applications in robotics or computer vision. The Kinect v1 (Microsoft) release in November 2010 promoted the use of RGB-D cameras, so that a second version of the sensor arrived on the market in July 2014. Since it is possible to obtain point clouds of an observed scene with a high frequency, one could imagine applying this type of sensors to answer to the need for 3D acquisition. However, due to the technology involved, some questions have to be considered such as, for example, the suitability and accuracy of RGB-D cameras for close range 3D modeling. In that way, the quality of the acquired data represents a major axis. In this paper, the use of a recent Kinect v2 sensor to reconstruct small objects in three dimensions has been investigated. To achieve this goal, a survey of the sensor characteristics as well as a calibration approach are relevant. After an accuracy assessment of the produced models, the benefits and drawbacks of Kinect v2 compared to the first version of the sensor and then to photogrammetry are discussed.

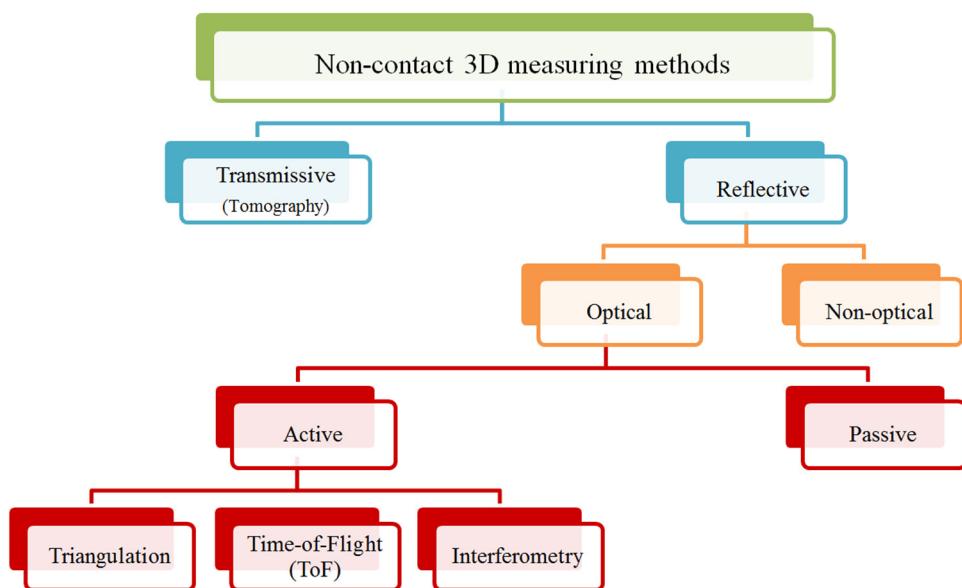
**Keywords:** RGB-D camera; Kinect v2; computer vision; calibration; 3D modeling; accuracy assessment

---

## 1. Introduction

Saving three-dimensional information about geometry of objects or scenes tends to be increasingly applied in the conventional workflow for documentation and analysis, of cultural heritage and archaeological objects or sites, for example. In this particular field of study, the needs in terms of restoration, conservation, digital documentation, reconstruction or museum exhibitions can be mentioned [1,2]. The digitization process is nowadays greatly simplified thanks to several techniques available that provide 3D data [3]. In the case of large spaces or objects, terrestrial laser scanners (TLS) are preferred because this technology allows collecting a large amount of accurate data very quickly. While trying to reduce costs and working on smaller pieces, on the contrary, digital cameras are commonly used. They have the advantage of being rather easy to use, through image-based 3D reconstruction techniques [4]. Besides, both methodologies can also be merged in order to overcome their respective limitations and to provide more complete models [5,6].

In this context, regarding some aspects like price or computation time, RGB-D cameras offer new possibilities for the modeling of complex structures, such as indoor environments [7]. Indeed, these sensors enable acquiring a scene in real-time with its corresponding colorimetric information. Among them, the Kinect sensor developed by Microsoft in 2010 and the Asus Xtion Pro in 2011, based on the PrimeSense technology [8,9], have encountered a great success in developer and scientific communities. Since July 2014, a second version of Microsoft's Kinect sensor has been available, based on another measurement technique. Such range imaging devices make use of optical properties, and are referred to as active sensors since they use their own light source for the active illumination of the scene. However, whereas Kinect v1 was based on active triangulation through the projection of structured light (speckle pattern), Kinect v2 sensor has been designed as a time-of-flight (ToF) camera [10]. Thanks to this technology, a depth measurement to the nearest objects is provided for every single pixel of the acquired depth maps. To illustrate this change of distance measurement principle, Figure 1 offers an overview of non-contact 3D measuring methods.



**Figure 1.** Time-of-Flight technology among non-contact 3D measuring methods. A more detailed taxonomy of active vision techniques can be found in [11].

In the following sections, an overview of application fields and calibration methods related to range imaging devices is presented. Then, characteristics of the recent Kinect v2 sensor are summarized, through a set of investigations carried out according to different aspects. The specific calibration process is also reported, in order to analyse its influence on the metric performances of the camera. Once all these parameters are known, the question of close range 3D modeling approaches can be addressed. To make a conclusion on the ability of Kinect v2 to provide reliable 3D models, the accuracy of the produced data has to be assessed. Considering not only the accuracy obtained but also the computation time or ease of use of the proposed methodology, finally a discussion about improvements and possible use of the sensor is reported.

## 2. Related Works

Because of their attractiveness and imaging capacities, lots of works have been dedicated to RGB-D cameras during the last decade. The aim of this section is to outline the state-of-the-art related to this technology, considering aspects such as fields of application, calibration methods or metrological approaches.

### 2.1. Fields of Application of RGB-D Cameras

A wide range of applications can be explored while considering RGB-D cameras. The main advantages are the cost, which is low for most of them compared to laser scanners, but also their high portability which enables a use on board of mobile platforms. Moreover, due to a powerful combination of 3D and colorimetric information as well as a high framerate, RGB-D cameras represent an attractive sensor well-suited for applications notably in the robotic field. Simultaneous localization and mapping (SLAM) tasks [12] or dense 3D mapping of indoor environments [13,14] are good examples of this problematic. Besides, industrial applications can be mentioned for real-time change detections [15] or for detections on automotive systems [16]. Some forensics studies are also reported, for crime scene documentation [17].

The Kinect sensors marketed by Microsoft are motion sensing devices, initially designed for gaming and entertainment purposes. They enable a contactless interaction between the user and a games console by gesturing. Since the release of the Kinect v1 sensor in 2010, lots of studies involving the device in alternative applications have been released. The capacities of localization [18] and SLAM systems [19] with Kinect devices have also been investigated. Moreover, pose recognition and estimation of human body [20] or hand articulations [21,22] have been considered. To answer to the need for 3D printing, [23] suggests the design of a scanning system based on a Kinect-type consumer-grade 3D camera. Finally, with the more recent Kinect v2 sensor, applications for face tracking [24], coastal mapping [25] or cyclical deformations analysis [26] appear in the literature.

### 2.2. Towards 3D Modeling of Objects with a RGB-D Camera

The creation of 3D models represents a common and interesting solution for the documentation and visualization of heritage and archaeological materials. Because of its remarkable results and its affordability, the probably most used technique by the archaeological community remains

photogrammetry. However, the emergence in the past decades of active scanning devices to acquire close-range 3D models has provided new possibilities [27]. Given their characteristics, RGB-D cameras belong to the most recent generation of sensors that could be investigated for such applications. The original use of Kinect v1 device has already been modified for the creation of a hand-held 3D scanner [28] or for a virtual reality project [29]. Since the sensor can also provide 3D information in the form of point clouds, the geometrical quality of data acquired on small objects such as a statue [30], or on bigger scenes such as archaeological walls [31] has also been assessed. More recently, the use of a Kinect v2 sensor with the Kinect Fusion tool [32] has shown good results for the direct reconstruction of 3D meshes [33].

### 2.3. Error Sources and Calibration Methods

The main problem while working with ToF cameras is due to the fact that the measurements realized are distorted by several phenomena. For guarantying the reliability of the acquired point clouds, especially for an accurate 3D modeling purpose, a prior removal of these distortions must be carried out. To do that, a good knowledge of the multiple error sources that affect the measurements is useful. A description of these sources of measuring errors is summarized in [34,35]. Considering especially the Kinect gaming sensors, analysis of related error sources are reported for example in [36,37].

First of all, a systematic depth-related deformation, also depicted as systematic wiggling error by some authors [38] can be reported. This deformation is partially due to inhomogeneities within the modulation process of the optical beam, which is not perfectly sinusoidal. Lots of works have been devoted to the understanding of this particular error. As a matter of fact, calibration of time-of-flight sensors was a major issue in many studies in the last decade. Most of the time, the introduced approaches are based either on Look-Up Tables [39] to store and interpolate the related deviations, or on curve approximation with B-splines [38] to model the distance deviations. In other methods of curve approximation, polynomials have been used [40] to fit the deviations. These models require a smaller number of initial values than B-splines, but are also less representative of the actual deviations. Whatever the method, the aim of this calibration step is the storage of depth residuals as a function of the measured distance.

Since object reflectivity varies with the range and can also cause a distance shift depending on the distance to the observed object, an intensity-related (or amplitude-related) error must be mentioned. Specific calibration steps for its correction have been investigated, for example on a PMD time-of-flight camera [41]. Furthermore, the depth calibration needs to be extended to the whole sensor array, because a time delay of the signal propagation is observed as a function of the pixel position on the sensor array. A per-pixel distance calibration [42] can be considered. Besides, [39] suggests the computation of a Fixed Pattern Noise (FPN) matrix containing an individual deviation value for each pixel, that allows the diminution of the global surface deformations. Sometimes, this global correction is performed together with the wiggling error correction, by considering in the same mathematical model the depth-related error and the location of the pixel on the array [40]. Finally, some non-systematic depth deformations which rather correspond to noise are also reported. One should notice the existence of denoising and filtering methods for their reduction [43], as well as methods related to the multiple returns issue [44].

### 3. Survey of Kinect v2 Specifications

Since Kinect v2 is initially a motion sensing device produced to be used with a gaming console, it seems obvious that the measurements provided by it will be affected by some unavoidable error sources. Indeed, the environment in which the acquisitions are performed has an influence (e.g., temperature, brightness, humidity), as well as the characteristics of the observed object (reflectivity or texture, among others). Once the camera is on its tripod, the user intervention is limited to choosing settings (time interval, types of output data), and therefore, the automation level seems to be relatively high. However, errors can occur due to the internal workings of the sensor itself.

To quantify the accuracy of the acquired data, a good knowledge of these potential sources of errors is required. For the Kinect v1, some of these aspects have been investigated, such as pre-heating time [9], or the influence of radiometry and ambient light [45]. Because of the change of the distance measurement principle in the Kinect v2, the phenomena observed with this new sensor might be different. In spite of this change, it is important to investigate their influence on the produced data. After a review of the way the sensor works and the data it provides, a few experimentations are presented in this section. They deal with pre-heating time, noise reduction, and some environment or object-related criteria.

#### 3.1. Sensor Characteristics and Data Acquisition

Kinect v2 sensor is composed of two cameras, namely a RGB and an infrared (IR) camera. The active illumination of the observed scene is insured by three IR projectors. Some features of the sensor are summarized in Table 1. For example, given the specifications, at 2 m range a pixel size of 5 mm can be obtained.

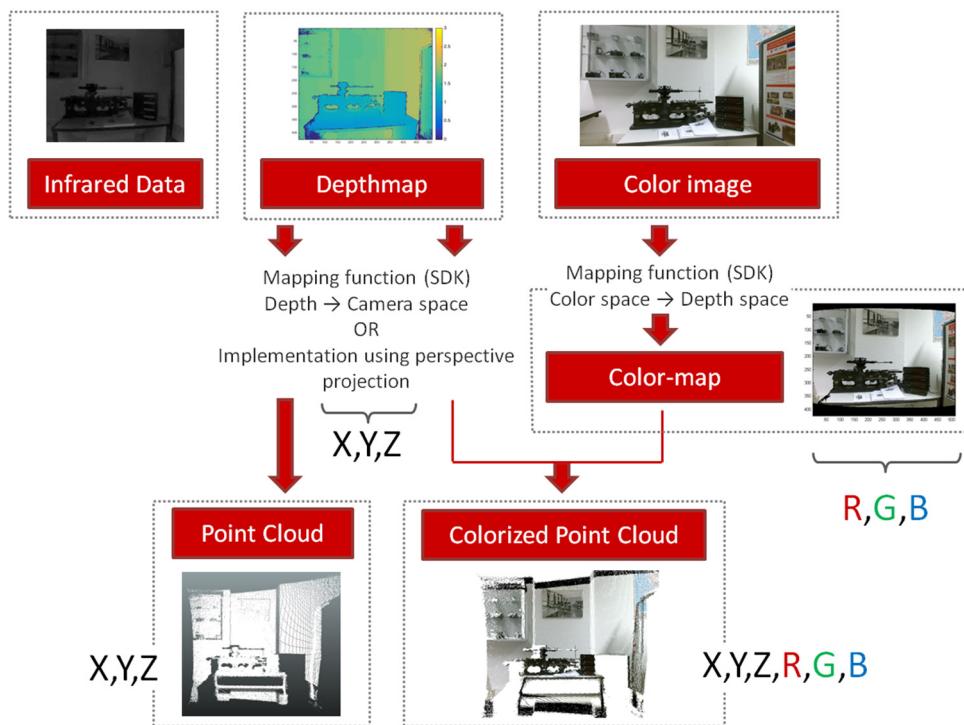
As mentioned on the Microsoft website [46], the Kinect v2 sensor is based on ToF principle. Even though time-of-flight range imaging is a quite recent technology, many books deal with its principles and its applications [47–50]. The basic principle is as follows: knowing the speed of light, the distance to be measured is proportional to the time needed by the active illumination source to travel from emitter to target. Thus, matricial ToF cameras enable the acquisition of a distance-to-object measurement, for each pixel of its output data. It should be noted that, unlike other ToF cameras, it is impossible to act on the modulation frequency or within the integration time of the input parameters of the Kinect v2 sensor.

**Table 1.** Technical features of Kinect v2 sensor.

<b>Infrared (IR) camera resolution</b>	512 × 424 pixels
<b>RGB camera resolution</b>	1920 × 1080 pixels
<b>Field of view</b>	70 × 60 degrees
<b>Framerate</b>	30 frames per second
<b>Operative measuring range</b>	from 0.5 to 4.5 m
<b>Object pixel size (GSD)</b>	between 1.4 mm (@ 0.5 m range) and 12 mm (@ 4.5 m range)

As shown in Figure 2, three different output streams arise from the two lenses of the Kinect v2 device: infrared data and depthmaps come from one lens and have the same resolution. The depthmaps are 2D images 16 bits encoded in which measurement information is stored for each pixel. The color images of higher resolution come from the second lens. The point cloud is calculated thanks to the depthmap, because of the distance measurements it contains. From depth values stored in the pixel matrix, two possible ways can be considered to infer 3D coordinates from the 2D data. Either a mapping function provided by the SDK (Software Development Kit) is applied between depth and camera space, or self-implemented solutions can be used applying the perspective projection relationships. The result is a list of (X, Y, Z) coordinates that can be displayed as a point cloud.

If colorimetric information is required, a transformation of the color frame has to be performed because of its higher resolution compared to the depthmap. This transformation is achieved based on a mapping function from the SDK, which enables to map the corresponding color to a corresponding pixel in depth space. The three dimensional coordinates coming from the depthmap are combined with the corresponding color information (as RGB values) and constitute a colorized point cloud.

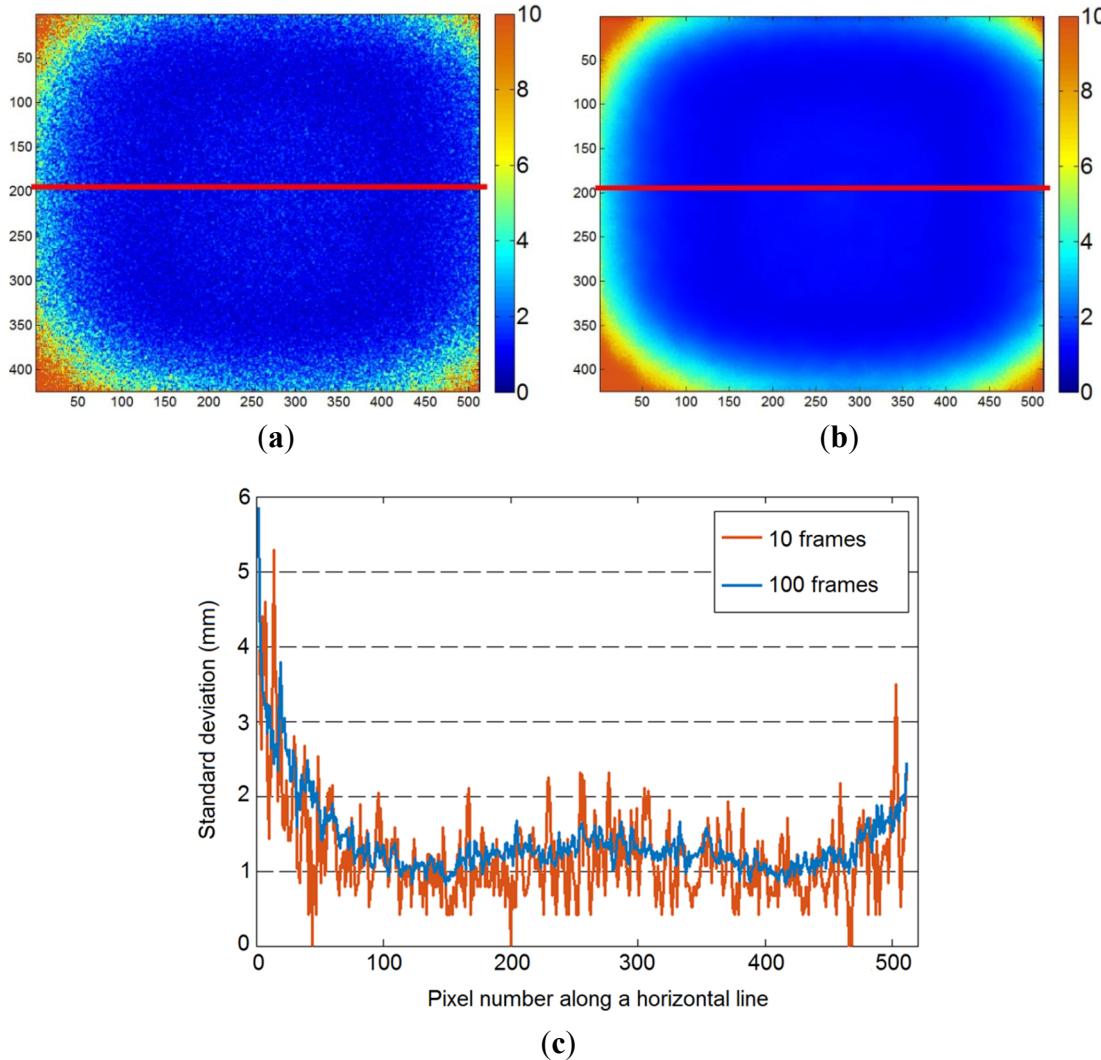


**Figure 2.** Schematic representation of the output data of Kinect v2 and summary of point cloud computation.

### 3.2. Pre-Heating Time and Noise Reduction

The first study involved the necessary pre-heating time. Indeed, as shown in [9,51], some RGB-D sensors need a time delay before providing reliable range measurements. In order to determine this time delay which is necessary for the Kinect v2 to reach constant measurement, the sensor was placed parallel to a white planar wall at a range of about 1.36 m. Three hundred and sixty depthmaps were recorded during one and a half hours, which is one depthmap recorded each 15 seconds. An area of  $10 \times 10$  pixels corresponding to the maximal intensities in IR data was then selected in each recorded

depthmap. Indeed, the sensor is not placed perfectly parallel to the wall, thus it cannot be assumed that the shortest measured distances are located at the center of the depthmaps. Mean measured distances can be calculated and represented as a function of the operating time. The distance varies from 5 mm up to 30 minutes and becomes then almost constant, around 1 mm. It can be noticed that the sensor's cooling fan starts after 20 minutes. For future tests, the pre-heating time of 30 minutes will be respected even if the distance variation is quite low. This experiment is related in [52].



**Figure 3.** Visualization of color-coded standard deviation calculated for each pixel over (a) 10 successive frames, and (b) 100 successive frames; (c) Observation of standard deviations along a line of pixels (see red line) in the set of 10 frames (orange) and 100 frames (blue).

Secondly, the principle of averaging several depthmaps has also been investigated in order to reduce the measurement noise inherent to the sensor and its technology [45]. With this process, the influence of the frame number on the final measurement precision is studied. For this purpose, the sensor was placed in front of a planar wall, at a distance of 1.15 m. To investigate the influence of the frames averaging, different sizes of samples have been considered. Figure 3a,b presents the standard deviation computed for each pixel on datasets of respectively 10 successive frames and 100 successive frames.

The standard deviation values obtained with 10 frames are lower than 1 cm, except on the corners. Considering the averaging of 100 successive frames, these values are not lower, but a smoothing effect is observed especially on the corners. The variation of the standard deviations along a horizontal line of pixels in the middle of the frame (see red line) is presented in Figure 3c, where the case of 10 frames and 100 frames are plotted together. Regarding this graph it is clearly visible that there is no gain in terms of precision when more frames are averaged, because the mean standard deviation values are almost the same in both cases. Only the variations of the standard deviations are reduced for the higher number of frames. The smoothing effect observed while increasing the number of frames appears through the diminution of the peaks in Figure 3c, expressing a reduction of the noise. In conclusion, the use of a larger sample does not really enhance the precision but it makes it more uniform. That is why for the future acquisitions, datasets of about 10–50 frames will be considered. Besides, similar results obtained for the case of real scenes are reported in [52].

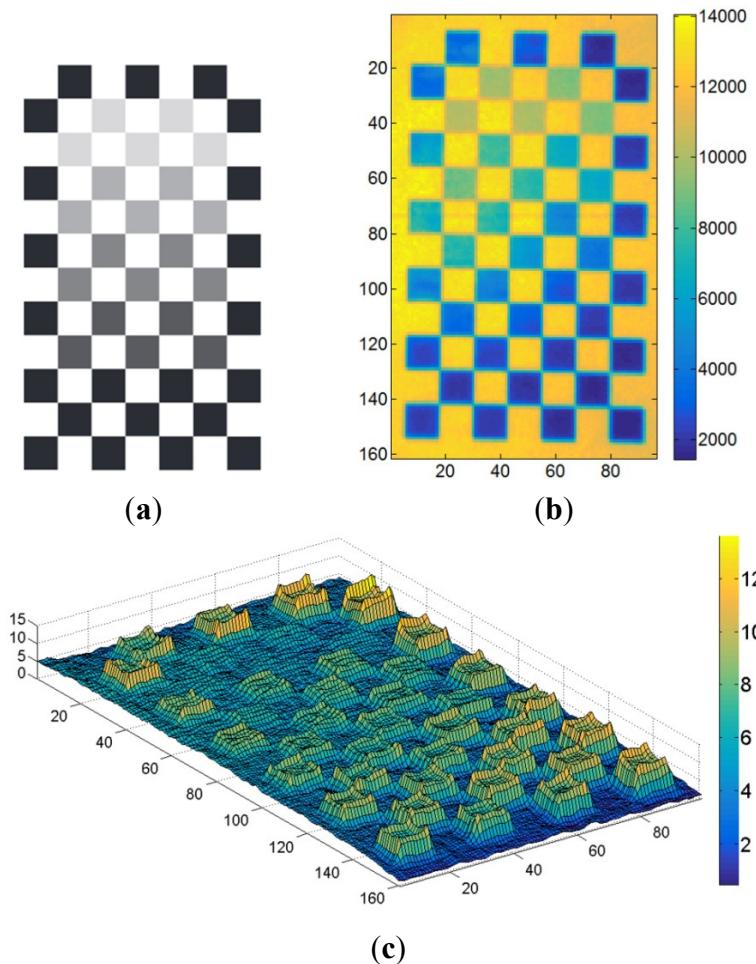
### 3.3. Influences of the Environment and Acquired Scene

The previous section described phenomena which are rather related to the sensor itself. However, some characteristics of the environment in which the acquisitions are realized also have an influence on the measured distances.

#### 3.3.1. Color and Texture of the Observed Items

The effects of different materials on intensity as well as on distance measurements have first been assessed using samples characterized by different reflectivities and roughness. The sensor was placed parallel to a board with samples. It appears that reflective as well as dark materials stand out among all the samples. Indeed, very reflective as well as very dark surfaces display intensity values which are lower than for other samples. In the depthmaps, this phenomenon results in measured distances which are larger than expected. The more important effect is observed for a compact disk, *i.e.*, a highly reflective material. For this sample, distance measurements vary up to 6 cm in an inhomogeneous way. Always considering very reflective materials, experiments performed on mirror surfaces have confirmed the previous findings with measurement deviations that can reach up to 15 cm.

To complete these investigations, observations have been realized on a checkerboard of grey levels inspired from [38] (Figure 4). As illustrated in Figure 4b, the displayed intensity increases with the brightness of the patterns. Intensity values displayed by black squares are almost six times lower than the ones displayed by white squares (2000 against 12,000). Besides, considering the values stored in the depthmaps (Figure 4c), one can conclude that the lower the intensities are, the longer the measured distances are. Deviations of the distance up to 1.2 cm are observed for the black patterns. As mentioned in Subsection 2.3, the reduction of intensity-related errors can be considered in a specific calibration process described in [41] for a PMD ToF camera. This correction technique is not discussed here. Considering the modeling approach developed in Section 5 of this paper, the reconstructed object does not show significant color changes. Hence, in that case this error source does not really affect the measurements and is safely negligible.

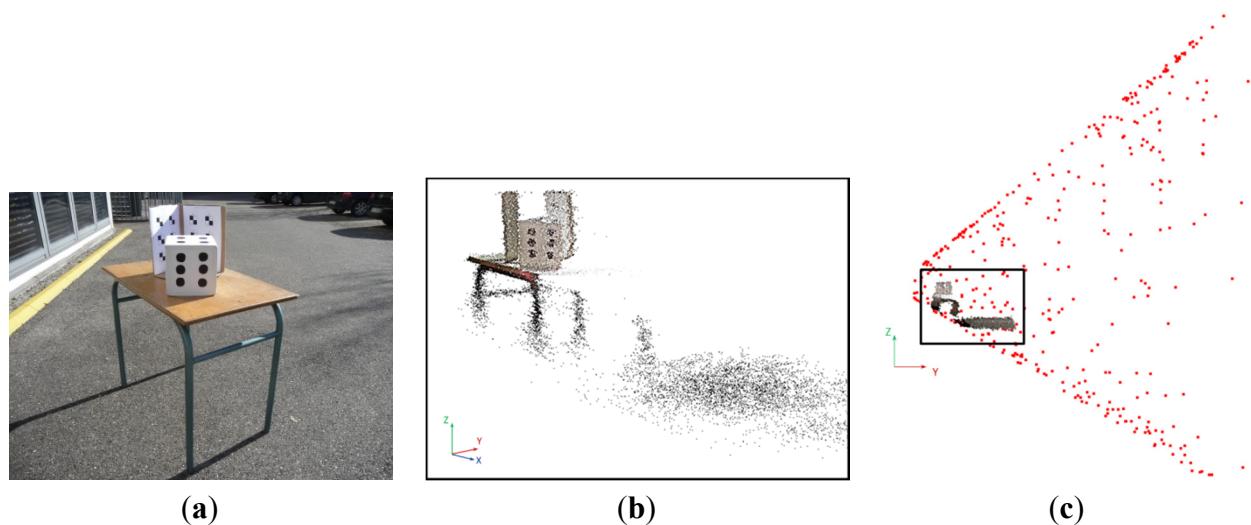


**Figure 4.** (a) Checkerboard of grey levels inspired from [38], (b) intensity values measured on this checkerboard, and (c) corresponding 3D representation of distance measurements variations (in mm).

### 3.3.2. Behavior in High Brightness Conditions

Because the previous Kinect device was not suited for sunny outdoor acquisitions, the influence of the brightness conditions on the measurements realized with a Kinect v2 was finally studied. Some applications of the sensor could require a proved efficiency in outdoor conditions. Hence, acquisitions were performed with the sensor during a sunny day. The observed scene composed of a table and a squared object is presented on Figure 5a, as well as the resulting point cloud on Figure 5b. It appears that parts of the scene are missing, such as the table stands. As a matter of fact, about 35,000 points are visible on this point cloud over the 217,088 maximal points of a single point cloud, which is about 20%. Regarding Figure 5c which represents the entire point cloud, about 2% of “flying pixels” (in red on Figure 5c) are included in the total number of points.

Despite a clear lack of points to depict the whole scene, the results show that the sensor is able to work during a sunny day provided that the light does not directly illuminate the sensor. Indeed, strong backlighting conditions cause the sensor’s disconnections from the computer. Two phenomena summarize the previous observations: the number of “flying pixels” is clearly visible particularly on the edges of the sensor field of view, and the number of acquired points decreases when the light intensity raises.



**Figure 5.** Outdoor acquisitions: (a) Picture of the observed scene; (b) Corresponding point cloud acquired with the sensor, (c) and the same entire point cloud in a profile view (scene in a black rectangle) without removal of the “flying pixels” (red).

#### 4. Calibration Method

A major drawback of 3D ToF cameras is the significant number of errors that can influence the acquired data. These errors, which affect both the overall performances of the system and its metrological performances, were emphasized in the related work section as well as in the section related to the survey of the sensor.

This section deals essentially with the assessment and the correction of the systematic distance measurement errors. To do that, a calibration method divided into several steps is proposed. Considering the fact that range imaging cameras combine two technologies, two types of calibrations have to be considered. Firstly, the geometric calibration of the lens is performed in order to correct its geometric distortions. Secondly, a depth calibration is suggested in order to assess and correct depth-related errors.

##### 4.1. Geometric Calibration

Since lenses are involved in the Kinect device acquisitions, geometric distortions as reported with DSLR cameras can be observed. For the first version of the Kinect sensor, some computer vision based calibration algorithms have been developed [53,54]. Unfortunately, these algorithms have not yet been adapted for the Kinect v2 sensor. Nevertheless, it appears in lots of works, e.g., [55], that time-of-flight cameras can be geometrically calibrated with standard methods.

As for common 2D sensors, several images of a planar checkerboard have been taken under different points of view. It is worth noting that the infrared data have been used to handle this geometric calibration, because depth and infrared output streams result from the same lens. To determine the necessary intrinsic parameters, our dataset was treated with the “Camera Calibration Toolbox” proposed by [56] under the Matlab software. In this tool, the Brown distortion model is implemented. This model is largely used for photogrammetric applications because of its efficiency.

The “Camera Calibration Toolbox” allows the determination of radial as well as tangential distortion parameters, and also internal parameters such as focal length and principal point coordinates.

With this technique applied, one should underline the fact that changing or removing some of the images causes the computed results to vary by a few pixels. This phenomenon is only due to the low sensor resolution. As a matter of fact, the best calibration results are obtained regarding the lower uncertainties calculated on the parameters. An overview of results obtained for these parameters is available in [52], as well as in [57]. The direct integration of the distortion coefficients into our point clouds computation algorithm allows the obtainment of a calibrated three dimensional dataset, which is no longer affected by the actual distortions of the initial depthmap.

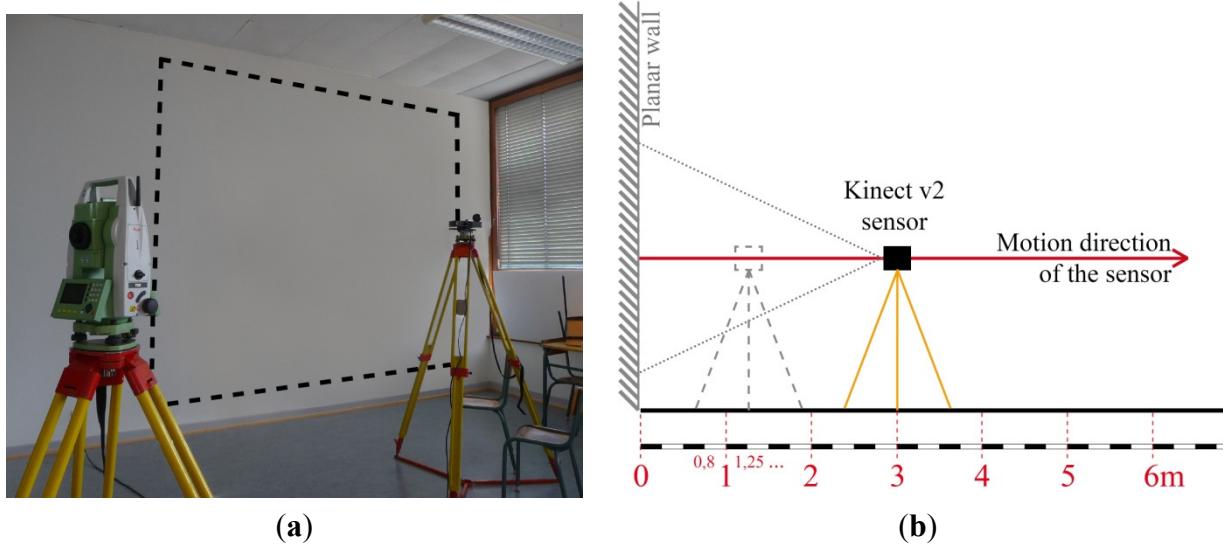
#### 4.2. Depth Calibration

As mentioned in Section 2.3, errors related to the measurement principle systematically affect the distances stored in the depth images. Accordingly, a depth-related calibration model needs to be defined to correct the measured distances and thus enhance the quality of the acquired data. This step is particularly significant since the aim of this study is to produce 3D models in an as accurate as possible way. After a description of the method that has been set up for this purpose, the processing of the data and the results that they provide are analyzed in this section.

##### 4.2.1. Experimental Setup

A common way to assess the distance inhomogeneity consists on positioning the camera parallel to a white planar wall at different well-known ranges. In our study, the wall has been surveyed beforehand with a terrestrial laser scanner (FARO Focus<sup>3D</sup>, precision 2 mm). The standard deviation computed on plane adjustment in the TLS point cloud is 0.5 mm. It allows confirming the reference wall planarity assuming that the TLS provides higher accuracy than the investigated sensor. After that, a line has been implanted perpendicularly to the wall by tachometry. Conventional approaches make use of track lines [39,42] to control the reference distances between plane and sensor, or rather try to estimate the camera position with respect to the reference plane through a prior calibration of an additional combined CCD camera [38,40]. In our experimental setup, marks have been determined by tachometry along the previously implanted line at predetermined ranges. They represent stations from which the acquisitions are realized with the Kinect sensor placed on a static tripod. A picture of the setup is shown in Figure 6a.

Stations have been implanted at 0.8 m range from the reference wall, and then from a 1 m up to 6 m range every 25 cm (Figure 6b). In this way, the sensor is progressively moved away from the plane, while covering the whole operating range of the camera. To limit the influence of noise, at each sensor position, 50 successive depthmaps of the wall have been acquired with intervals of one second. Finally, to insure the parallelism between sensor and wall and thus to avoid the addition of a possible rotation effect, the distances have been surveyed by tachometry at each new position, using small prisms at the two sensor extremities.



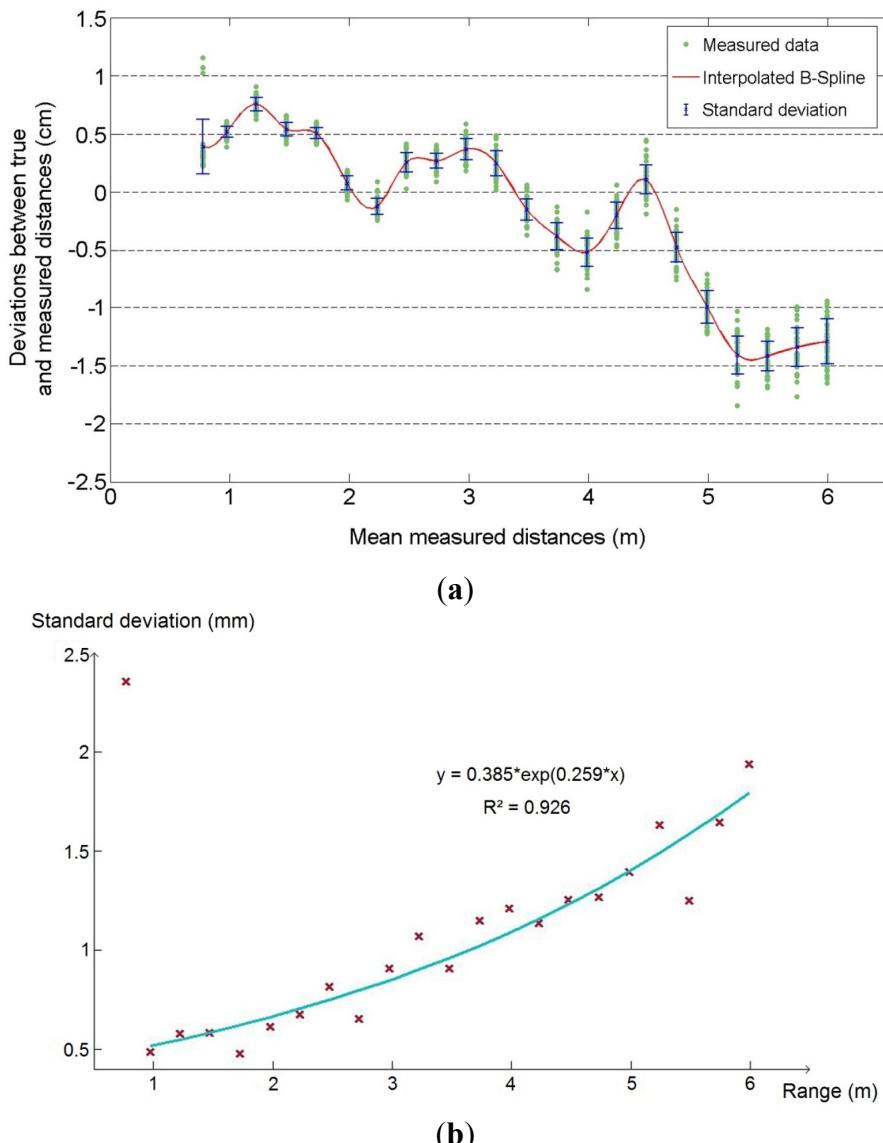
**Figure 6.** (a) Picture of the experimental setup with the observed area in dotted lines; (b) Global schema of the acquisition protocol.

#### 4.2.2. Data Analysis

The results obtained led us to consider two steps of depth calibration. First, deviations of measurements for a small central patch of the frames are assessed through B-spline fitting as suggested in [42]. Secondly, the deformations extended to the whole sensor array are investigated, in order to highlight the need of a potential per-pixel correction.

First of all, a central matrix of  $10 \times 10$  pixels is considered in the input images acquired during the experiment. This enables to compute mean measured distances from the sensor for each position. Then, the deviations between real and measured distances are plotted on a graph as a function of the range. Each of the 50 deviations obtained from the 50 depthmaps acquired per station is represented as a point. As depicted in Figure 7a, a B-spline function is estimated within these values. Since the sensor was accurately placed on the tripod with respect to its fixing screw, a systematic offset occurs on raw measurements because the reference point for the measurement does not correspond to the optical center of the lens. The influence of this offset corresponding to the constant distance between fixing point and lens (approximately 2 cm) is removed on this graph. It appears that the distortions for the averaged central area vary from  $-1.5$  cm to  $7$  mm, which is rather low regarding the technology investigated. At  $4.5$  m range, a substantial variation is observed. Under  $4.5$  m range, the deviations are rather included within an interval of variation of almost  $1$  cm (from  $-1.5$  cm to  $7$  mm).

Since a set of 50 successive depthmaps is acquired for each position of the sensor, a standard deviation can also be computed over each sample. Figure 7b presents a separate graph showing the evolution of the computed standard deviations as a function of the range. As it can be seen, the standard deviation increases with the range. This means that the scattering of the measurements increases around the mean estimated distance when the sensor moves away from the scene. Moreover, for the nearest range ( $0.8$  m), the standard deviation reported stands out among all other positions. As a matter of fact, measurements realized at the minimal announced range of  $0.5$  m would probably be still less precise. Since a clear degradation with depth is showed, it makes sense to bring a correction.

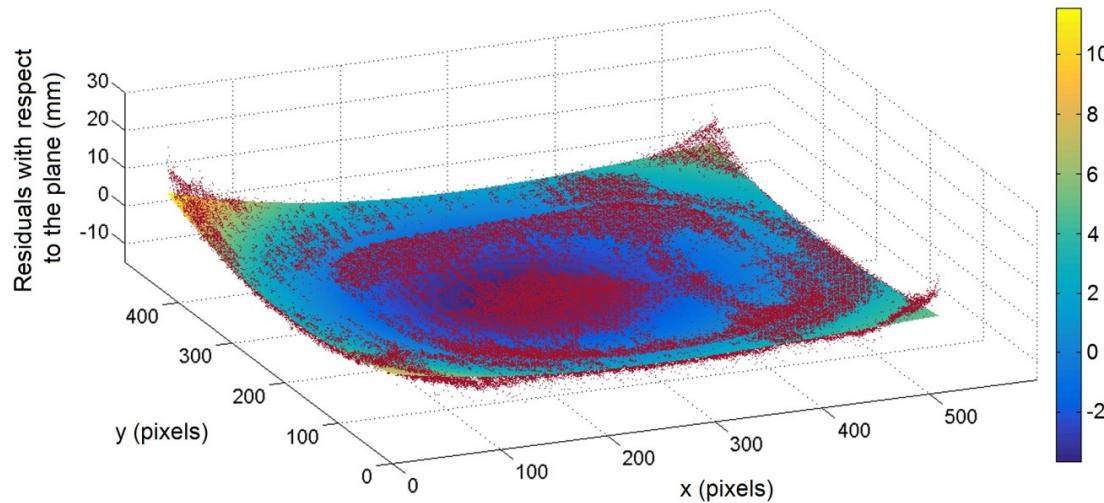


**Figure 7.** (a) Deviations (in cm) between true distances and measured distances, as a function of the range (m). (b) Evolution of the standard deviation (in mm) calculated over each sample of 50 measurements, as a function of the range (m).

#### 4.2.3. Survey of Local Deformations on the Whole Sensor Array

The second part of our depth calibration approach consists of extending the deformations analysis to the whole sensor array. For this purpose, only the first 10 datasets acquired on the planar wall corresponding to the first 10 positions from 0.8 to 3 m are considered. It appears that the corresponding point clouds are not planar, but rather exhibits a curved shape. On these 10 successive point clouds, planes have been fitted by means of least-squares adjustment. To realize a proper adjustment in the Kinect point clouds, some outlying points corresponding to null values in the depthmaps have to be previously removed. Then, the Euclidean distances between a point cloud and its corresponding plane are calculated. It provides a matrix of residuals, in which for each pixel the deviation of the actual measured distance with respect to the adjusted plane is stored.

Figure 8 shows the residuals with respect to the plane at 1 m range. The residuals have been computed and represented in the form of a point cloud (in red), in which a surface is interpolated (here color-coded). First, it appears that the residuals are more important at sensor boundaries and especially on the corners. As known from the literature, these radial effects increase with the distance between sensor and target. In our case, the residuals can reach about 1 centimeter up to tens of centimeters in the corners at 3 m range.



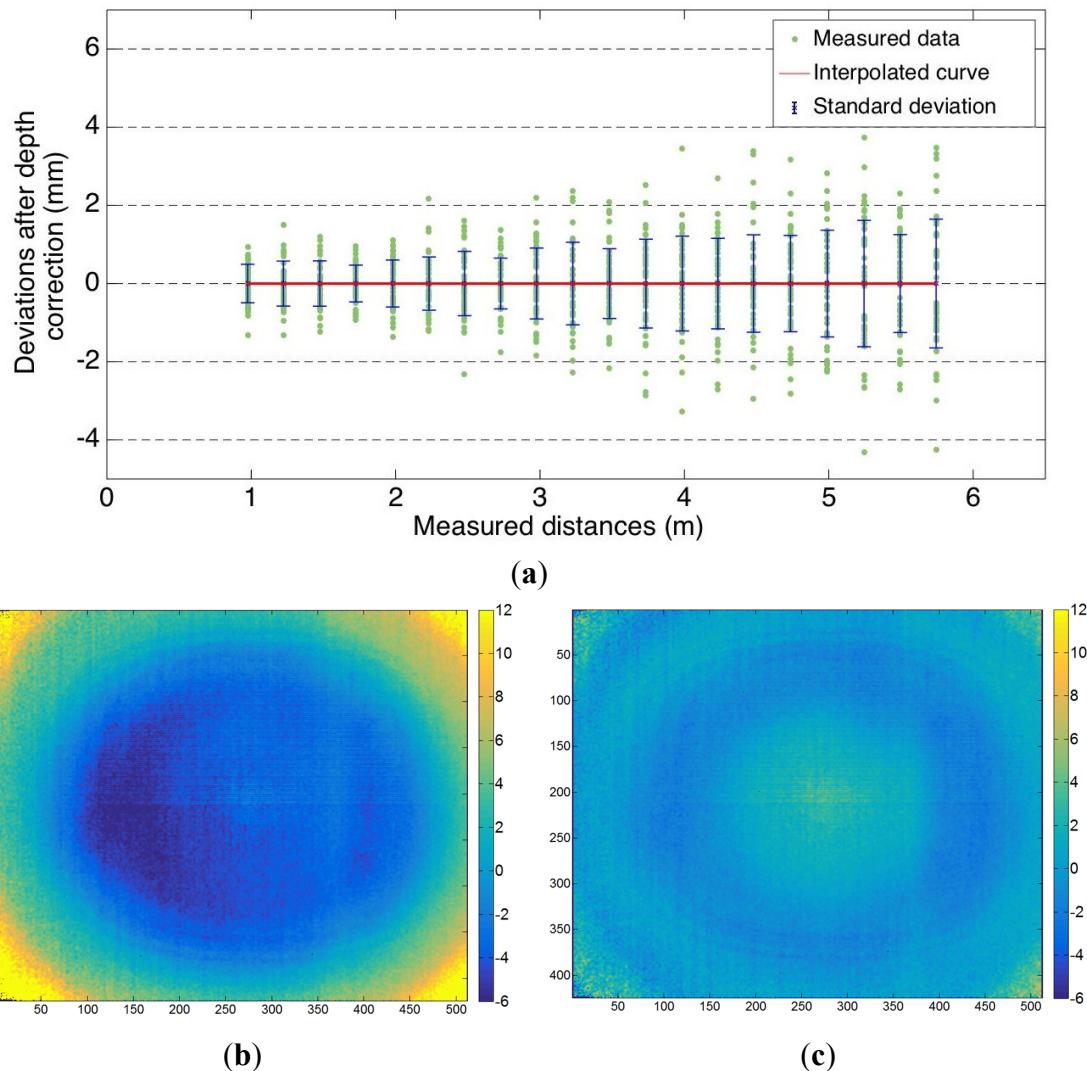
**Figure 8.** Point cloud of residuals (mm) in red, at 0.8 m, together with the surface (color-coded) fitting the residuals.

To define an adapted correction, the parameters of the interpolated surfaces previously mentioned are stored. The chosen surface model is polynomial, in order to keep a convenient computation time for the correction step. This post-processing step consists of delivering signed corrections interpolated from the residuals surfaces computed at each range. Thus, to correct each pixel of the depthmap, given the distance value it contains, a residual value will be interpolated considering the nearest ranges. In that way, a correction matrix is computed, that needs to be added to the initial depthmap as a per-pixel correction.

#### 4.3. Influence of Corrections

All the corrections related to depth calibration are part of post-processing carried out on the depthmaps. In order to visualize the influence of the computed correction on the systematic distance measurement error, the dataset consisting of measurements realized on a planar wall at known ranges is used. To apply the calibration, an adjustment of the errors model in the form of a spline previously determined is required. Indeed, knowing the curve parameters and the distance measured allows inferring the correction values applied to the measurements. For the analysis, a central patch of  $10 \times 10$  pixels on the corrected data is considered as before. The remaining deviations between real and corrected distances are shown in Figure 9a. It appears that the average deviations are reduced to zero through this calibration. These good results can also be explained by the fact that the dataset investigated is the one used to determine the correction parameters. However, it enables validating the model. At the same time, the standard deviations computed over the 50 corrected measurements per

station are almost the same or have slightly decreased compared to the initial data. They confirm the increase of the scattering of the measurements when the range raises, but they look promising for the investigated device.



**Figure 9.** (a) Remaining deviations (in mm) between real and corrected distances, as a function of the range (m); Colorized representation of the residuals (in mm) with respect to a fitted plane (range: 1.25 m), (b) before correction and (c) after correction of local deformations.

Considering the correction extended to the whole sensor array, the approach used is based on surface approximation within the residual values calculated with respect to a plane. These residuals are visible on a colorized representation for the initial data (Figure 9b) and after correction of the local deformations (Figure 9c) with the method proposed beforehand. A 1.25 m range is considered in the figures. Whereas before correction the residuals have negative values in a circular central area and can reach more than 1 cm on the corners, they are reduced to some millimeters after correction. The residuals on the corners are also much less significant, so that a positive impact of the second step of the calibration is noticed. For the 10 ranges considered for this correction, the residuals are always reduced after calibration. This is also true for the standard deviation calculated on plane fitting residuals. For the ranges considered, the

standard deviations vary from about 2–5 mm before correction, and from about 1–3 mm after correction. Nevertheless, a less significant effect is observed at 2.25 m range. In addition, these corrections are time-consuming due to the local processing of each pixel.

Consequently, depending on the accuracy required by the applications, the need for a depth calibration must be discussed. Indeed, the user could decide to use only the measurements provided by a limited central area of the sensor, regarding the fact that fewer deformations affect the middle of the sensor array. However, the works presented in the following section of this paper have all been realized with calibrated data.

## 5. Experimental Approach for 3D Modeling with Kinect v2

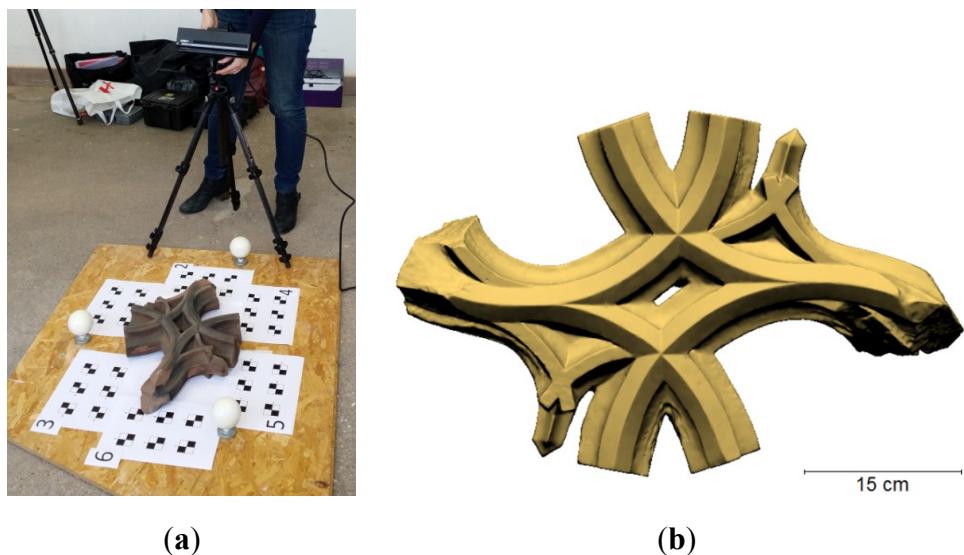
To perform a three-dimensional reconstruction of an object or a scene with the Kinect sensor, several ways can be considered for the alignment and the merge of the raw data coming from the sensor. Among them, the SDK provided by Microsoft offers a tool called “Kinect Fusion” [32] to answer to a possible use of the Kinect sensor as a 3D scanning device. In the literature, many works have shown the potential of Kinect v1 for modeling purposes. In [30], a standard deviation of about 1 mm is reached for the 3D reconstruction of a small statue. A precision in the order of magnitude of some millimeters is also reported in [58] for spheres and planes adjustment using acquisitions from a Kinect v1 sensor. Because of the technical differences between first and second version of Kinect, an improvement in terms of precision can be expected.

Regarding the calibration results obtained in this article, a priori error lower than 5 mm can be assumed for acquisitions realized at ranges smaller than 4 m. Even if a sub-millimetric precision can be offered by laser scanner or photogrammetry techniques for 3D reconstruction of small objects, it is interesting to test the Kinect v2 sensor for similar applications. In this section, a reconstruction approach with this device is discussed after the presentation of the object under study. This will lead to a qualitative assessment of the models produced.

### 5.1. Object under Study and Reference Data

The object under study is a fragment of a sandstone balustrade of about  $40 \times 20$  centimeters, coming from the Strasbourg Cathedral Notre-Dame (France). As shown in Figure 10a, it presents almost a symmetric geometry and flat faces, except on its extremities.

A measuring arm from FARO (FARO ScanArm) was used to generate a ground truth dataset of the fragment. It provides a very high metric quality thanks to its sub-millimetric accuracy (0.04 mm according to the specifications). Such a device is often used for metrological purposes, but because of its cost, its bad portability and its fragility, other devices are often preferred in many missions. In this study, the dataset obtained with the measuring arm constitutes a reference and will be used to assess the quality of the model created with the Kinect v2 sensor. The reference point cloud of the fragment contains about 5 million points after a spatial resampling of 0.1 mm. Based on this point cloud, a meshed model was created using the commercial software 3D Reshaper (Technodigit). The resulting mesh is composed of about 250,000 faces (Figure 10b).



**Figure 10.** (a) Object under study (a sandstone balustrade fragment); and (b) reference mesh realized after acquisition with a measuring arm.

## 5.2. Reconstruction Approach

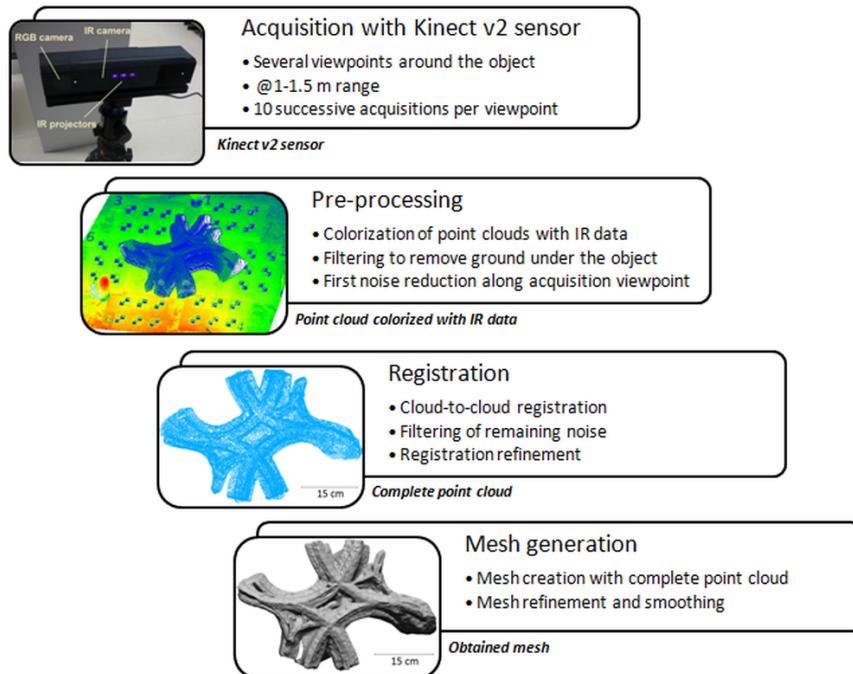
In the following subsections, the experimental workflow used to reconstruct the object under study with a Kinect v2 sensor is described. Once the data have been acquired, the produced point clouds need to be registered before the final creation of a meshed model.

### 5.2.1. Data Acquisition and Registration

First of all, a set of data of the balustrade fragment has been acquired with the Kinect v2 sensor. The way used to carry out the acquisitions is rather inspired by the common workflow set up in image-based reconstruction methods. A circular network composed of acquisitions from eight different viewpoints was performed around the object under study. An overlap of about 80% on the object was reported between two successive range images. In order to overcome the deformations affecting the measurement on the borders of the sensor, the object was located in the central part of the sensor. As shown in Figure 10a, flat targets were also placed around the object during the acquisitions for the point clouds' registration to be done in a post processing step. The sensor was placed at about 1 m of the object. Indeed, the results of depth calibration presented in the previous section emphasize the fact that considering scene observation at a range from 1–4 m, the global depth-related deformations are smaller. At the chosen range, it is possible to capture the whole object within each range image, as well as a sufficient number of flat targets for the future registration. At this range, the related spatial resolution of the object is about 2–6 mm, depending on the shapes encountered.

The survey of the sensor specifications in Section 3.2 also shows that the frame averaging allows a reduction of the measurement noise inherent to the sensor. Thus, 10 successive acquisitions were realized at the framerate of 0.5 s for each viewpoint. Based on raw data, the averaging of the depth images acquired for each viewpoint as well as the computation of individual point clouds from these depth images are performed. Then, a target-based registration was performed by extracting the target centers manually. A spatial resolution between 2 and 4 mm is reported on the ground where the targets

are located. This method based on target selection provides an overall registration error of about 2 cm, even with a refinement of the registration using the ICP method. This is high regarding the requirements for 3D modeling and the dimensions of the fragment, especially since the more adverse effect appears in the center of the frames where the object is located.



**Figure 11.** Processing chain applied in this study for 3D reconstruction based on Kinect, from the acquisition of point clouds to the mesh creation.

Thus, another method has been considered for point clouds registration. The workflow applied is summarized in Figure 11. The first step consists in the colorization of the point clouds with the infrared data provided by the sensor. This allows removing the ground under the object in the point clouds, through filtering. Moreover, noise reduction has been performed depending on the point of view. Then, homologous points have been selected directly on the object on pair-wise successive point clouds. It appears that this method is more reliable than the previous one. Indeed, after a registration refinement, the residual registration error between two successive point clouds varies from 2–4 mm. This operation has been repeated to register the eight individual point clouds which are counted to be between 13,000 and 14,000 points each. These better results are due to the fact that the common points chosen directly on the object are hence located close to the center of the point cloud. The use of targets directly on the object could be considered.

### 5.2.2. Creation of a Meshed Model

The complete point cloud of the balustrade fragment, as well as the mesh based on it and created through the 3D Reshaper software, are presented in Figure 11. Noise remains in the obtained complete point cloud and is hard to remove without losing geometric information. Consequently, the creation of the mesh is time-consuming. Indeed, several parameters can be empirically set during the mesh creation, which have a substantial influence on the resulting mesh. They can provide, for example, a good visual

smoothed result but with a loss of details. Because of the low quality of the initial point cloud, even the processing steps' order might influence the resulting mesh obtained following this methodology.

As depicted in Figure 11, the geometry of the final mesh suffers from several faults compared to the real fragment. Its metric accuracy will be assessed in the next Subsection. There is obviously a correlation between the deformations observed on the final mesh and error sources related to the raw data themselves. First of all, border effects can be observed along the line of sight of the sensor and constitute noise. A second problem is related to deformations of the geometry linked to the viewpoint. Among all error sources, these two problems which were detailed in [33] obviously contribute to the low visual quality of the mesh performed.

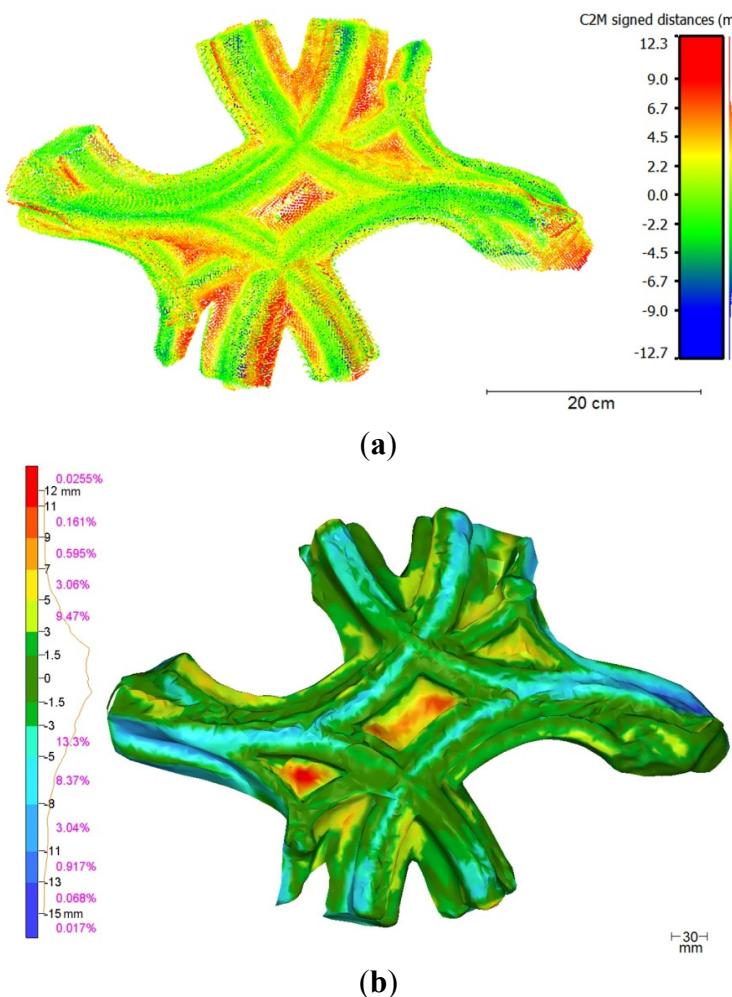
### 5.3. Accuracy Assessment

In order to assess the accuracy of the Kinect-based model, the reference mesh previously presented was used in order to make a series of comparisons between them. The reference point cloud obtained by the measuring arm offers a very high density of points which is almost 15 times larger than the density of the point cloud obtained with the Kinect v2. This high point density variation may be an issue for the cloud-to-cloud comparison algorithms, resulting in a loss of efficiency in the distance computation. Additionally, considering meshes, a loss of geometric quality or a data simplification due to mesh creation can occur. Comparisons between different kinds of data were therefore carried out.

The complete point cloud of the balustrade fragment and its corresponding mesh derived from Kinect raw data were compared to the reference mesh. Figure 12a presents the results of the cloud-mesh comparison and Figure 12b presents the results of the comparison between meshes. The deviations between datasets are color-coded, and projected on the computed data. Despite a possible bias due to the triangulation method applied during the mesh creation, for both methods the results look similar.

In both comparisons, a mean deviation of about 2 mm is reported for almost 40% of the models, and a standard deviation of about 2.5 mm is observed. Besides, maximal deviations of about 1.5 cm are reached. They correspond to local deformations affecting the model and are particularly high on the borders of the object. These deformations of at least 1 cm provide to the object false shapes, especially for the edges on the upper part of the object, or on its extremities.

Regarding the calibration results (Figure 7a), for the used range an *a priori* error in the order of magnitude of 5 mm can be assumed. It appears that 90% of the deviations are lower than the *a priori* error. Thus, the final error of about 2 mm on a large part of the model is more than acceptable for the device under study. It looks promising since a lot of improvements in the processing chain can be considered. Nevertheless, the remaining deviations in the order of magnitude of one centimeter still represent a significant percentage of the global size of the object. Besides, the visual rendering suffers from many deformations and the model is too far from the real shape of the object. This might be detrimental to archeological interpretation.



**Figure 12.** (a) Comparison between Kinect-based point cloud and reference mesh; and (b) comparison between mesh created from Kinect point cloud and reference mesh. The deviations are in millimeters.

#### 5.4. Kinect Fusion Tool

A first version of Kinect Fusion tool was already proposed for the first version of the device, and it has been adapted and directly added to the SDK of Kinect v2. It allows creating automatically accurate meshes of objects with a visual rendering close to reality. It is mainly based on an Iterative Closest Point (ICP) algorithm [59], a well-known algorithm in the field of computer vision.

In [58], the benefits of using Kinect Fusion tool compared to individual scans of a Kinect v1 are exposed in terms of accuracy and precision. For example, at 2 m range a precision of about 1 mm on sphere estimation is reached thanks to Kinect Fusion, whereas the precision is larger than 7 mm with the sensor alone. Despite the common use of this 3D reconstruction tool for Kinect sensors due to their good results, the major drawback of this tool is the “black box” effect related to the use of the standalone version available in the SDK, for which the algorithms are not available. Indeed, the user can only change a few settings but do not really manage all processing steps that are carried out on the raw data. However, improvements of the workflow applied in Kinect Fusion are possible through open source implementations and libraries. This was, for example, already tested with Kinect v1 in [60],

where they suggest adding calibration results and estimated exterior parameters to the implementation. The deviations in the models produced are reduced compared to the initial Kinect Fusion models.

In [33], the Kinect Fusion tool associated with the Kinect v2 sensor was used in its standalone version to scan another architectonic fragment. After a cloud-mesh comparison with a reference model, a standard deviation of 1.4 mm and a mean deviation of 1.1 mm were reported for the Kinect Fusion model. Since this fragment was more complex than the one investigated in this article, this led us to conclude about the satisfactory precision provided by Kinect Fusion. Moreover, the aim of this article was rather to analyze a more experimental methodology; that is why the Kinect Fusion tool will not be further investigated here.

## 6. Discussion

This section aims at analyzing the results obtained for the 3D model creation with the Kinect v2 sensor. Not only will the previously assessed accuracy be commented on, but rather the characteristics of the experimental reconstruction method will presented along with potential improvements. First of all, some differences between Kinect v1 and the v2 investigated in this article will be discussed.

### 6.1. Differences between Kinect v1 and v2

The Kinect v2 sensor presents several technical benefits compared to the first version. First of all, it offers a larger horizontal as well as vertical field of view. According to this criterion and to the change of depth sensing technology, the point clouds obtained present a better resolution. Moreover, the point clouds produced by Kinect v1 sensor show striped irregularities, which is not the case with the second version. Then, the resolution of color images are increased. To summarize these technical features, specifications of both Kinect versions can be found in the form of tables in previous contributions as for example [24] or [37]. Finally, as depicted in Section 3.3.2, outdoor acquisitions can be envisaged since Kinect v2 seems to be much less sensitive to daylight.

Considering now the accuracy of the data captured by both versions of the sensor, a metrological comparison between Kinect v1 and Kinect v2 carried out in [37] is examined. The assessment of data acquired on a same artefact shows that accuracy and precision are better with Kinect v2. There is also less influence of the range on the decrease in precision for Kinect v2 measurements. Data performed for other applications as 3D reconstruction also confirm the benefits of the second version compared to the first one. In the field of people tracking, for example, it is shown in [61] that the tracking is 20% more accurate with Kinect v2 and it is also possible in outdoor environments. It appears that Kinect v2 is generally better in many cases. However, Kinect v1 and v2 both suffer from a limited accuracy on edges.

### 6.2. RGB-D Camera as an Alternative to Photogrammetry?

Since a 3D reconstruction method for close-range objects with Kinect v2 has been presented in this paper, it seems interesting to compare it to a more common approach. For this purpose, [33] presents models obtained on the same archaeological material through DSLR camera pictures acquisition and use of an image-based modeling solution. With a mean deviation smaller than 1 mm and very local small deviations, photogrammetry obviously provides a high level of detail on very accurate models.

As expected, the results obtained in Section 5 of this paper are clearly still far from achieving such accuracy. Nevertheless, regarding the initial use of the Kinect v2 device, the accuracy cannot constitute a unique comparison point. To extend the comparison to further considerations, other characteristics should be investigated. Table 2 offers a comparative overview between both mentioned methodologies. Not only the accuracy reached has been taken into account, but also criteria such as acquisition and computation times, or cost, have been considered in order to underline the benefits and drawbacks of each strategy.

Whereas the cost is comparable for the digital camera and the Kinect sensor, it seems that the photogrammetric survey remains easier to apply than the method considered in Section 5. This is also true in terms of portability because of the external electricity supply still required by the Kinect sensor. Besides, the more experimental method consisting of registration of Kinect point clouds remains very time-consuming for a final lower accuracy. Indeed, it appeared that in the meshes presented in the previous section that a lot of details were not at all or badly represented, which is associated with a loss of geometrical information. These results tend to confirm the higher efficiency of image-based reconstruction methods when accurate 3D models are required, regardless of the object size.

**Table 2.** Comparative overview of some criteria for the image-based and for the Kinect-based 3D reconstruction methods.

	Photogrammetry (DSLR Camera)	Kinect v2 (Registration Method)
Cost	low to medium	low
Portability	good	good, but requires USB 3 and external electricity supply
Object size	from small objects to entire rooms or buildings	rather limited to objects
Indoor/outdoor	indoor + outdoor	mainly indoor
Ease-of-use	rather easy, needs care for photographs acquisition	needs some practice, for registration especially
Processing time	some hours all in all	long, because of manual treatments
Accuracy	very high, up to sub-millimetric (depends on the camera)	≈1 cm, to be improved

### 6.3. Suggested Improvements

Regarding the results obtained for the mesh creation in Section 5.3, it is obvious that the Kinect sensor presents technical constraints which are actually problematic for 3D modeling applications. Nevertheless, improvements in the reconstruction method are necessary in order to enhance the metrical quality of the models produced. Since there is no unique approach for aligning and merging Kinect raw data in order to get a three-dimensional reconstruction of a scene, several solutions can be considered. An approach would be to use the colorimetric information produced by the camera [31]. In the methodology developed in the abovementioned contribution, SURF algorithm [62] was used for the localization and matching of keypoints on color images obtained from a Kinect v1 device. After that, a first alignment is provided, which can be refined by using the ICP principle. This idea of combining the RGB information as a complement to frame-to-frame alignment based on 3D point clouds helps overcoming the reduced accuracy of the depth data.

Other open issues related to the considered acquisition process and registration method have to be investigated. Indeed, it has been shown in [33] that some geometric deformations are directly related to the acquisition viewpoints. As a matter of fact, the choice of a network of eight circular acquisitions realized in the proposed experiment is questionable. The presence of almost one more zenithal view could perhaps influence the final result. Besides, depending on the geometry of the object observed, the idea of performing stripes rather than circular networks should also modify the result. Regarding the registration methodology carried out, its process was very long because of the manual tasks required. A higher automation of the registration could result in a reduction in processing time, but also a removal of potential user handling influences. Moreover, the pair-wise alignment of successive point clouds presents the problem of error propagation, so that a global adjustment could be considered to minimize the global error. Another approach to consider would be to apply weights to the points selected for the registration, as a function of the incident angle of the point clouds they belong to. This would confer more importance to more reliable measurements, probably enhancing at the same time the quality of the registration.

## 7. Conclusion

This article offers an overview of a set of experiments realized with a RGB-D camera, in order to assess its potential for close-range 3D modeling applications. The sensor studied is a low-cost motion sensing device from Microsoft, namely the Kinect v2. The main idea was to investigate its measurement performance in terms of accuracy, as well as its capacity to be used for a particular metrological task. To do that, experiments for raw data enhancement and a calibration approach were presented. An experimental method for 3D reconstruction with Kinect was also described, with its related issues.

Several tests contributed to highlighting errors arising from the environment and the properties of the captured scene, as well as errors related to the sensor itself. As reported in many contributions dealing with RGB-D cameras, a pre-heating time has to be considered to obtain constant measurements. It has been estimated that almost 30 minutes is needed with the Kinect v2 sensor. Then, acquiring a tenth of successive depthmaps from a single viewpoint enables reducing the measurement noise through temporal averaging of the measured distances. Besides, it appeared that the color and reflectivity of the items have a great influence, both on the intensity images and on the distance measurements provided by the camera. Considering outdoor efficiency, the achieved results look promising.

Geometric as well as depth calibrations have been performed. Whereas the geometric lens distortions can be computed with a standard method also used for DSLR cameras, depth calibration requires further studies. A systematic depth-related error reported for range imaging devices was first reduced. The remaining deviations of some millimeters after calibration are very satisfactory. Finally, the local deformations which appear especially on the sensor array boundaries were reduced. This quite time-consuming step does not really enhance the measurements observed for the central area of the sensor array. Nevertheless, it was applied for the acquisitions dedicated to 3D modeling.

An experimental methodology for 3D reconstruction of an archaeological fragment was finally reported. Based on a mainly manual workflow from point cloud registration to mesh creation, it was proved to be very time-consuming. Besides, regarding the accuracy achieved while using the same

sensor with the automatic tool Kinect Fusion, the models presented in Section 5 are obviously highly correlated to that explored in this paper. As a matter of fact, even if their accuracy in an order of magnitude of 1 cm is rather satisfactory, many improvements could be considered. For example, an acquisition process including more zenithal views could enable reducing the influence of deformations related to grazing angles. Adding colorimetric information is also a possible improvement, as well as more automation during the registration process.

Regarding the small price difference between RGB-D cameras and photogrammetric survey, as well as the wide range of robust software available for photogrammetry, as expected this second solution remains the easiest and most accurate method for 3D modeling issues. However, one should underline that the use of a Kinect sensor for this purpose was a real challenge. However, the technology of RGB-D camera looks promising for many other applications, once the major limitations in terms of geometric performances have been addressed. This kind of devices seems to have a bright future, and this can be confirmed when considering, for example, the Project Tango [63] developed by Google. A smartphone or a tablet including a range imaging camera is used to reconstruct a scene in real-time, so that it looks attractive, easy and fast. Similarly, the Structure Sensor [64] offers a 3D scanning solution with a tablet, based on structured light projection and detection similarly to the first Kinect version. Since data quality assessment is still an open issue in such projects, acquisitions performed with such innovative techniques will no doubt be the focus of future contributions.

## Acknowledgments

The authors would like to thank the Oeuvre Notre-Dame foundation and the City of Strasbourg for the provision of the Cathedral's balustrade fragment. They also want to thank Marie-Anne Mittet for her advices particularly concerning the setting up of the device under study.

## Author Contributions

Elise Lachat realized its Master thesis project entitled “Assessment of a 3D camera (Kinect v2) for measurement and reconstruction of close-range 3D models” in the Photogrammetry and Geomatics group of the INSA of Strasbourg, under the supervision of Tania Landes and Pierre Grussenmeyer. This article is based on the results obtained during this project. The experiments were realized with the help of Hélène Macher, who also contributed to the writing of this paper. Tania Landes and Pierre Grussenmeyer also participated to the writing of this paper.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Remondino, F. Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sens.* **2011**, *3*, 1104–1138.

2. Remondino, F.; Rizzi, A.; Agugiaro, G.; Girardi, S.; De Amicis, R.; Magliocchetti, D.; Girardi, G.; Baratti, G. Geomatics and geoinformatics for digital 3D documentation, fruition and valorization of cultural heritage. In Proceedings of the EUROMED 2010 Workshop “Museum Futures: Emerging Technological and Social Paradigms”, Lemesos, Cyprus, 8–13 November 2010.
3. Sansoni, G.; Trebeschi, M.; Docchio, F. State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors* **2009**, *9*, 568–601.
4. Hullo, J.F.; Grussenmeyer, P.; Fares, S. Photogrammetry and dense stereo matching approach applied to the documentation of the cultural heritage site of Kilwa (Saudi Arabia). In Proceedings of CIPA Symposium, Kyoto, Japan, 10–15 October 2009.
5. Grussenmeyer, P.; Alby, E.; Landes, T.; Koehl, M.; Guillemin, S.; Hullo, J.F.; Smigiel, E. Recording approach of heritage sites based on merging point clouds from high resolution photogrammetry and terrestrial laser scanning. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *XXXIX-B5*, 553–558.
6. El-Hakim, S.F.; Beraldin, J.A.; Picard, M.; Godin, G. Detailed 3D reconstruction of large-scale heritage sites with integrated techniques. *IEEE Comput. Graphic. Appl.* **2004**, *24*, 21–29.
7. 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. *Int. J. Robot. Res.* **2012**, *31*, 647–663.
8. Freedman, B.; Shpunt, A.; Machline, M.; Arieli, Y. Depth Mapping Using Projected Patterns. Patent US 2010/0118123, 13 May 2010.
9. Mittet, M.-A.; Grussenmeyer, P.; Landes, T.; Yang, Y.; Bernard, N. Mobile outdoor relative localization using calibrated RGB-D cameras. In Proceedings of the 8th International Symposium on Mobile Mapping Technology, Tainan, Taiwan, 1–3 May 2013; pp. 2–7.
10. Dal Mutto, C.; Zanuttigh, P.; Cortelazzo, G.M. *Time-of-Flight cameras and Microsoft Kinect™*; Springer Science & Business Media: New York, NY, USA, 2012.
11. Curless, B. Overview of active vision techniques. In Proceedings of SIGGRAPH 2000 Course on 3D Photography, New Orleans, LA, USA, 23–28 July 2000.
12. May, S.; Droeschel, D.; Holz, D.; Fuchs, S.; Malis, E.; Nüchter, A.; Hertzberg, J. Three-dimensional mapping with time-of-flight cameras. *J. Field Robot.* **2009**, *26*, 934–965.
13. Henry, P.; Krainin, M.; Herbst, E.; Ren, X.; Fox, D. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. In Proceedings of the 12th International Symposium on Experimental Robotics, Delhi, India, 18–21 December 2010.
14. Dong, H.; Figueroa, N.; El Saddik, A. Towards consistent reconstructions of indoor spaces based on 6D RGB-D odometry and KinectFusion. In Proceedings of the IEEE International Conference Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 1796–1803.
15. Kahn, S.; Bockholt, U.; Kuijper, A.; Fellner, D.W. Towards precise real-time 3D difference detection for industrial applications. *Comput. Ind.* **2013**, *64*, 1115–1128.
16. Hussmann, S.; Ringbeck, T.; Hagebeuker, B. A performance review of 3D TOF vision systems in comparison to stereo vision systems. In *Stereo Vision*; Bhatti, A., Ed.; InTech: Vienna, Austria, 2008; pp. 103–120.
17. Cavagnini, G.; Sansoni, G.; Trebeschi, M. Using 3D range cameras for crime scene documentation and legal medicine. *Proc. SPIE* **2009**, *7239*, doi:10.1117/12.806191.

18. Mittet, M.A.; Landes, T.; Grussenmeyer, P. Localization using RGB-D cameras orthoimages. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *XL-5*, 425–432.
19. Bakkay, M.C.; Arafa, M.; Zagrouba, E. Dense 3D SLAM in dynamic scenes using Kinect. In Proceedings of 7th Iberian Conference on Pattern Recognition and Image Analysis, Santiago de Compostela, Spain, 17–19 June 2015; pp. 121–129.
20. Shotton, J.; Sharp, T.; Kipman, A.; Fitzgibbon, A.; Finocchio, M.; Blake, A.; Moore, R. Real-time human pose recognition in parts from single depth images. *Commun. ACM* **2013**, *56*, 116–124.
21. Keskin, C.; Kıracı, F.; Kara, Y.E.; Akarun, L. Real time hand pose estimation using depth sensors. In *Consumer Depth Cameras for Computer Vision*; Fossati, A., Gall, J., Grabner, H., Ren, X., Konolige, K., Eds.; Springer: London, UK, 2013; pp. 119–137.
22. Kuznetsova, A.; Leal-Taixé, L.; Rosenhahn, B. Real-time sign language recognition using a consumer depth camera. In Proceedings of 2013 IEEE International Conference on Computer Vision Workshops (ICCVW), Sydney, Australia, 2–8 December 2013; pp. 83–90.
23. Hsieh, C.T. A new Kinect-based scanning system and its application. *Appl. Mech. Mater.* **2015**, *764*, 1375–1379.
24. Amon, C.; Fuhrmann, F.; Graf, F. Evaluation of the spatial resolution accuracy of the face tracking system for Kinect for windows v1 and v2. In Proceedings of the 6th Congress of the Alps-Adria Acoustics Association, Graz, Austria, 16–17 October 2014.
25. Butkiewicz, T. Low-cost coastal mapping using Kinect v2 time-of-flight cameras. In Proceedings of 2014 Oceans-St. John's, St. John's, NL, Canada, 14–19 September 2014; pp. 1–9.
26. Lahamy, H.; Lichti, D.; El-Badry, M.; Qi, X.; Detchev, I.; Steward, J.; Moravvej, M. Evaluating the capability of time-of-flight cameras for accurately imaging a cyclically loaded beam. *Proc. SPIE* **2015**, *9528*, doi:10.1117/12.2184838.
27. Guidi, G.; Remondino, F.; Morlando, G.; Del Mastio, A.; Uccheddu, F.; Pelagotti, A. Performances evaluation of a low cost active sensor for cultural heritage documentation. In Proceedings of the 8th Conference on Optical 3D Measurement Techniques, Zurich, Switzerland, 9–12 July 2007.
28. UCSD Researchers Modify Kinect Gaming Device to Scan in 3D. Available online: [http://ucsdnews.ucsd.edu/archive/newsrel/general/2011\\_07arkinect.asp](http://ucsdnews.ucsd.edu/archive/newsrel/general/2011_07arkinect.asp) (accessed on 28 June 2015).
29. Richards-Rissetto, H.; Remondino, F.; Agugiaro, G.; Robertsson, J.; von Schwerin, J.; Girardi, G. Kinect and 3D GIS in archaeology. In Proceedings of 18th International Conference on Virtual Systems and Multimedia, Milan, Italy, 2–5 September 2012; pp. 331–337.
30. Menna, F.; Remondino, F.; Battisti, R.; Nocerino, E. Geometric investigation of a gaming active device. *Proc. SPIE* **2011**, *8085*, doi:10.1117/12.890070.
31. Quintana, M.; Zvietcovich, F.; Castaneda, B. 3D reconstruction of archaeological walls using Kinect. In Proceedings of the EUROMED 2014 Conference, Lemessos, Cyprus, 3–8 November 2014; pp. 47–58.
32. Newcombe, R.A.; Izadi, S.; Hilliges, O.; Molyneaux, D.; Kim, D.; Davison, A.J.; Fitzgibbon, A. KinectFusion: Real-time dense surface mapping and tracking. In Proceedings of 10th IEEE International Symposium on Mixed and Augmented Reality, Basel, Switzerland, 26–29 October 2011; pp. 127–136.
33. Lachat, E.; Macher, H.; Landes, T.; Grussenmeyer, P. Assessment of the accuracy of 3D models obtained with DSLR camera and Kinect v2. *Proc. SPIE* **2015**, *8528*, doi:10.1117/12.2184866.

34. Foix, S.; Alenya, G.; Torras, C. Lock-in time-of-flight (ToF) cameras: A survey. *Sens. J.* **2011**, *11*, 1917–1926.
35. Lefloch, D.; Nair, R.; Lenzen, F.; Schäfer, H.; Streeter, L.; Cree, M.J.; Kolb, A. Technical foundation and calibration methods for time-of-flight cameras. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*; Grzegorzek, M., Theobalt, C., Koch, R., Kolb, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 3–24.
36. Khoshelham, K. Accuracy analysis of Kinect depth data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2011**, *XXXVIII-5*, 133–138.
37. Gonzalez-Jorge, H.; Rodríguez-Gonzálvez, P.; Martínez-Sánchez, J.; González-Aguilera, D.; Arias, P.; Gestó, M.; Díaz-Vilarino, L. Metrological comparison between Kinect I and Kinect II sensors. *Measurement* **2015**, *70*, 21–26.
38. Lindner, M.; Schiller, I.; Kolb, A.; Koch, R. Time-of-flight sensor calibration for accurate range sensing. *Comput. Vis. Image Underst.* **2010**, *114*, 1318–1328.
39. Kahlmann, T.; Remondino, F.; Ingensand, H. Calibration for increased accuracy of the range imaging camera Swissranger™. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2006**, *XXXVI-5*, 136–141.
40. Schiller, I.; Beder, C.; Koch, R. Calibration of a PMD-camera using a planar calibration pattern together with a multi-camera setup. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2008**, *XXXVII-3a*, 297–302.
41. Lindner, M.; Kolb, A. Calibration of the intensity-related distance error of the PMD ToF-camera. *Proc. SPIE* **2007**, *6764*, doi:10.1117/12.752808.
42. Lindner, M.; Kolb, A. Lateral and depth calibration of pmd-distance sensors. In *Advances in Visual Computing: Second International Symposium, ISVC 2006 Lake Tahoe, NV, USA, November 6–8, 2006. Proceedings, Part II*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 524–533.
43. Lenzen, F.; Kim, K.I.; Schäfer, H.; Nair, R.; Meister, S.; Becker, F.; Theobalt, C. Denoising strategies for time-of-flight data. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*; Grzegorzek, M., Theobalt, C., Koch, R., Kolb, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8200, pp. 25–45.
44. Guomundsson, S.A.; Aanæs, H.; Larsen, R. Environmental effects on measurement uncertainties of time-of-flight cameras. In Proceedings of IEEE International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, Romania, 13–14 July 2007; Volume 1, pp. 1–4.
45. Chow, J.; Ang, K.; Lichti, D.; Teskey, W. Performance analysis of a low-cost triangulation-based 3D camera: Microsoft Kinect system. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *XXXIX-B5*, 175–180.
46. Kinect for Windows. Available online: <https://www.microsoft.com/en-us/kinectforwindows/default.aspx> (accessed on 28 June 2015).
47. Grzegorzek, M.; Theobalt, C.; Koch, R.; Kolb, A. *Time-of-Flight and Depth Imaging. Sensors, Algorithms and Applications*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8200.
48. Hansard, M.; Lee, S.; Choi, O.; Horaud, R.P. *Time-of-flight cameras: Principles, Methods and Applications*; Springer: London, UK, 2012.
49. Kolb, A.; Koch, R. *Dynamic 3D Imaging*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5742.
50. Remondino, F.; Stoppa, D. *TOF Range-Imaging Cameras*; Springer: Heidelberg, Germany, 2013.

51. Chiabrando, F.; Chiabrando, R.; Piatti, D.; Rinaudo, F. Sensors for 3D imaging: Metric evaluation and calibration of a CCD/CMOS time-of-flight camera. *Sensors* **2009**, *9*, 10080–10096.
52. Lachat, E.; Macher, H.; Mittet, M.A.; Landes, T.; Grussenmeyer, P. First experiences with Kinect v2 sensor for close range 3D modelling. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *XL-5/W4*, 93–100.
53. Herrera, C.; Kannala, J.; Heikkilä, J. Accurate and practical calibration of a depth and color camera pair. In Proceedings of 14th International Conference on Computer Analysis of Images and Patterns, Seville, Spain, 29–31 August 2011; pp. 437–445.
54. Kinect Calibration. Available online: <http://nicolas.burrus.name/index.php/Research/Kinect> (accessed on 28 June 2015).
55. Hansard, M.; Horaud, R.; Amat, M.; Evangelidis, G. Automatic detection of calibration grids in time-of-flight images. *Comput. Vis. Image Underst.* **2014**, *121*, 108–118.
56. Camera Calibration Toolbox for Matlab. Available online: [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/) (accessed on 28 June 2015).
57. Kim, C.; Yun, S.; Jung, S.W.; Won, C.S. Color and depth image correspondence for Kinect v2. In *Advanced Multimedia and Ubiquitous Engineering*; Park, J., Chao, H.C., Arabnia, H., Yen, N.Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; Volume 352, pp. 111–116.
58. Bueno, M.; Díaz-Vilarino, L.; Martínez-Sánchez, J.; González-Jorge, H.; Lorenzo, H.; Arias, P. Metrological evaluation of KinectFusion and its comparison with Microsoft Kinect sensor. *Measurement* **2015**, *73*, 137–145.
59. Besl, P.J.; McKay, N.D. Method for registration of 3-D shapes. *Proc. SPIE* **1992**, *1611*, doi:10.1117/12.57955.
60. Pagliari, D.; Menna, F.; Roncella, R.; Remondino, F.; Pinto, L. Kinect fusion improvement using depth camera calibration. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2014**, *XL-5*, 479–485.
61. Zennaro, S.; Munaro, M.; Milani, S.; Zanuttigh, P.; Bernardi, A.; Ghidoni, S.; Menegatti, E. Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications. In Proceedings of IEEE International Conference on Multimedia and Expo, Torino, Italy, 29 June–3 July 2015; pp. 1–6.
62. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In *Computer Vision–ECCV 2006*; Leonardis, A., Bischof, H., Pinz, A., Eds.; Springer: Berlin Heidelberg, Germany, 2006; Volume 3951, pp. 404–417.
63. Project Tango. Available online: <https://www.google.com/atap/project-tango/> (accessed on 28 June 2015).
64. Structure Sensor. Available online: <http://structure.io/> (accessed on 24 August 2015).