# A new marker-less 3D Kinect-based system for facial anthropometric measurements

Claudio Loconsole[1], Nuno Barbosa[2,3], Antonio Frisoli[1], and Verónica Costa Orvalho[2,4]

[1] PERCRO Laboratory, Scuola Superiore Sant'Anna
[2] Instituto de Telecomunicações
[3] Faculdade de Engenharia, Universidade do Porto
[4] Faculdade de Ciências, Universidade do Porto
{c.loconsole,a.frisoli}@sssup.it
nuno.barbosa@fe.up.pt,veronica.orvalho@dcc.fc.up.pt

**Abstract.** Several research fields like forensic and orthodontics, use facial anthropometric distances between a set of pairs of facial standard landmarks. There are tens of attempts to automatize the measurement process using 2D and 3D approaches. However, they still suffer of three main drawbacks: human manual intervention to fix the facial landmarks, physical markers placed on the face and required time for the measurement process.

In this paper, we propose a new marker-less system for automatic facial anthropometric measurements based on Microsoft Kinect and *FaceTracker* API. This new approach overcomes the three measurement drawbacks. We statistically validated the system with respect to the caliper-based manual system, through experimental measurements and one-way ANOVA test comparisons on 36 subjects. The achieved successful percentage in the comparison is equal to 54,5 %.

**Keywords:** facial anthropometry, depth camera, Kinect, measurement error, three-dimensional

## 1 Introduction

Facial[5] anthropometric measurements are commonly used in the following research fields:

- *forensic and physical anthropology* for finding the characterizing facial anthropometric measurements of populations. The traditional parameters considered in population studies are gender, age and geographical origin [2];
- *orthodontic, maxillofacial and speech researches* for treatment planning, preorthodontic and postorthodontic and/or surgical treatment and evaluation of postoperative swelling [3];

[5] According to [1], *"the face is the part of the front of the head between the ears and from the chin to the airline (or where it ought to be if you have lost it!). It includes the forehead, eyes, nose, mouth and chin."*

- *syndrome, paralysis and disease* for medical prevention and diagnosis purposes, both in adulthood and in childhood [4];
- *beauty and asymmetry* for investigation, modeling and improvement of the facial beauty [5];
- *face modeling and synthesizing* for Information Communication Technology targets (e.g. teleconference, entertainment, etc.) [6].

Usually, traditional methods of taking facial measurements are dependent on the competences and skills of the operator. In fact, the facial landmarks identification is performed manually by the operator, either by touching or by looking at the facial features. Furthermore, misalignment errors can arise due to the use of the caliper in the measurement of the 3D distances between the selected landmark pairs. Generally, we can identify three main drawbacks in the traditional measurement methods:

1. prerequisite training on live subjects (which can be painstaking);
2. the time-consuming nature of performing multiple direct measurements (especially when we require a large number of measurements);
3. the person to be measured must remain patiently still. For instance, a facial measurement of a child is harder to carry out than of an adult, due to the restlessness of the children.

Some attempts to use 3D data for automating the measurement process were considered in the works by *Kau et al.* [7] in 2005 and by *Hammond et al.* [8] in 2004. They use, respectively, high cost laser scanning and multiple camera systems and are still affected by the drawbacks mentioned above.

A more recent work uses a faster system based on IR cameras [9] that overcomes the cited drawbacks. But it also introduces the dependence on physical markers to be placed on the face to localize landmarks. The use of markers on a subject's face skin is intrusive for the subject itself and still requires human intervention to localize facial landmarks.

In this paper, we propose a new marker-less 3D Kinect-based system for facial anthropometric measurements using a single or multiple frontal shots of a human's face, providing the facial measurements. Our system overcomes the three drawbacks mentioned above, as well as the dependence on physical markers. It also lowers the cost for facial measurement in comparison with the solutions presented in [7] and in [8]. In order to validate the system as an automatic measuring approach, we also conducted an experimental test and statistical comparisons with standard measurement methods on 36 subjects. This research will contribute to all facial anthropometric research laboratories, speeding up and automating the process of facial measurement.

## 2 The proposed measurement method

The system we present provides a set of linear spatial measures that characterize the face of a subject. The measurements are made between pairs of standard

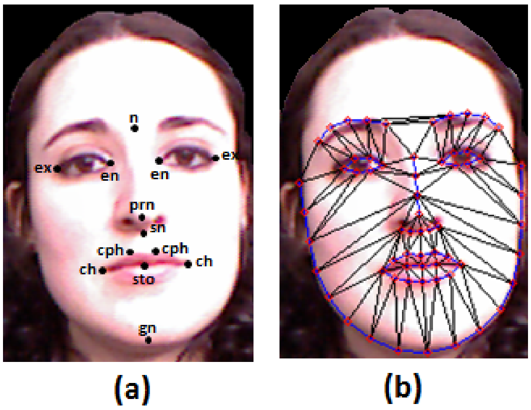


**Fig. 1.** The selected 13 facial landmarks for linear measurements of the face (a). The 66 facial landmarks identified and localized by *FaceTracker* API by *Saragih at al.* on the eye and oral-nasal regions and over the lower face contour (b).

facial landmarks. In Paragraph 2.1, we introduce and motivate the selection of the subset of facial landmarks and corresponding masurements used for our work. In Paragraph 2.2, we present the new system, implemented algorithms and their computational time performances.

### 2.1 Selected landmarks

*Weinberg et al.* [10] define 28 standard anthropometric facial landmarks and 19 standard linear measurements taken between selected pairs of these 28 landmarks. We use a subset of the *Weinberg et al.* standard landmarks and linear measurements. We selected 13 facial landmarks (see Fig. 1.(a) and Table 1) and 11 euclidean spatial measurements (see Table 1) that cover the eye and oral-nasal regions. Some standard facial landmarks, such as ear landmarks, have been excluded because they can not be correctly identified with frontal camera shots of the face. Other reasons for excluding landmarks are areas covered by hair, like eurion, and those requiring palpation, like gonion.

### 2.2 Description of the system and of the algorithms

Our system uses a Microsoft Kinect device (`http://www.xbox.com/kinect`, `http://www.primesense.com`) that provides visual and depth data, and a processing system. It uses OpenCV (`http://opencv.willowgarage.com`) OpenNI (`http://openni.org`) and PCL (`http://pointclouds.org`) open-source libraries. In order to avoid the use of physical markers or human interventions for landmark identification and localization, we employ a method by *Saragih et al.*. This method is based on constrained local models (CLM), optimized through a

**Table 1.** List of the selected subset of landmarks among standard anthropometric facial landmarks used by *Weinberg at al.* [10] (left). List of the selected subset of linear measurements between pairs of the selected landmarks (right). L stands for *left*, R stands for *right*.

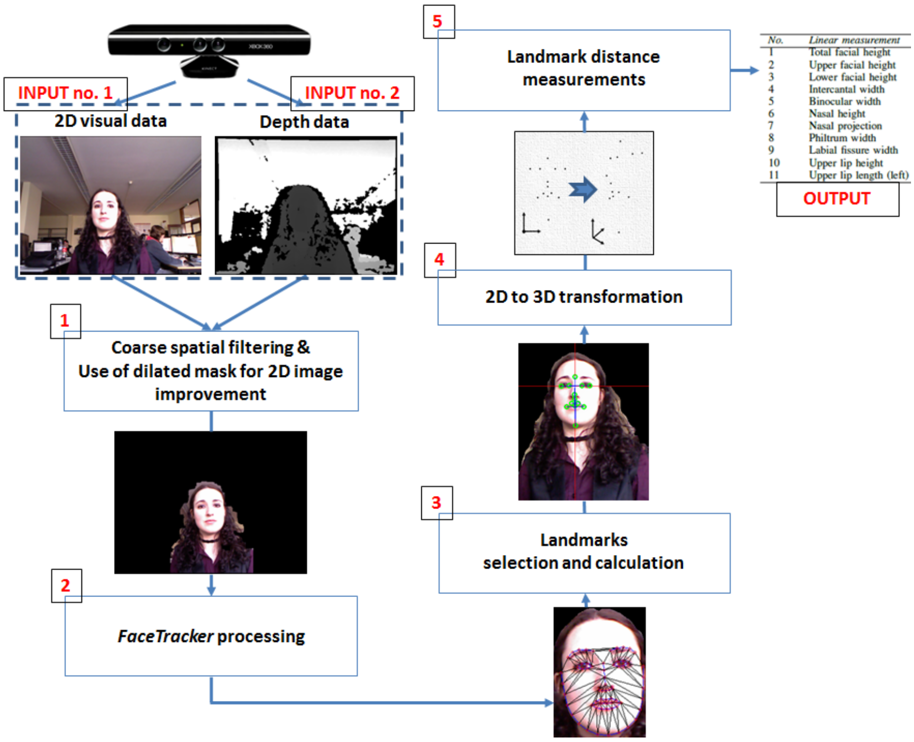
| *No.* | *Landmark* | *Description* | *No.* | *Linear measurement* | *Region* | *Landmarks* |
|---|---|---|---|---|---|---|
| 1 | n | nasion | 1 | Total facial height | Face | n-gn |
| 2 | gn | gnathion | 2 | Upper facial height | Face | n-sto |
| 3 | sto | stomion | 3 | Lower facial height | Face | sn-gn |
| 4 | sn | subnasale | 4 | Intercantal width | Eye | en-en |
| 5 | en (L) | left endocanthion | 5 | Binocular width | Eye | ex-ex |
| 6 | en (R) | right endocanthion | 6 | Nasal height | Nose | n-sn |
| 7 | ex (L) | left exocanthion | 7 | Nasal projection | Nose | sn-prn |
| 8 | ex (R) | right exocanthion | 8 | Philtrum width | Mouth | cph-cph |
| 9 | prn | pronasale | 9 | Labial fissure width | Mouth | ch-ch |
| 10 | cph (L) | left crista philtri inferior | 10 | Upper lip height | Mouth | sn-sto |
| 11 | cph (R) | right crista philtri inferior | 11 | Upper lip length (L) | Mouth | sn-ch |
| 12 | ch (L) | left chelion | | | | |
| 13 | ch (R) | right chelion | | | | |

homoscedastic kernel density estimate (KED) with an isotropic Gaussian kernel [11]. *Saragih et al.* provide a CLM-based C/C++ API for real time generic non-rigid face alignment and tracking called *FaceTracker* (`http://web.mac.com/jsaragih/FaceTracker/FaceTracker.html`). The *FaceTracker* allows the identification and localization of 66 2D landmarks on the face (see Fig. 1.(b)).

We consider a subset of these 66 2D landmarks composed by 14 points. Twelve out of 14 points are taken directly as seeds for 3D landmarks object of our study (Direct Landmarks: *gn, sto, sn, en (left), en (right), ex (left), ex (right), cph (left), cph (right), ch (left), ch(right)*, prn). The other two points are used to calculate, through average operation, the seed of the *n* landmark.

Using an offline preliminary calibration of the Microsoft Kinect, the intrinsic parameters of the Zhang camera model are estimated. Knowing these parameters and the depth data acquired by the Kinect, we can transform a 2D pixel point of the camera image into the corresponding 3D point in the space. So, the corresponding 3D points of the 2D landmarks can be directly calculated. Finally, the distances between selected 3D facial landmark pairs (see Table 1) can be calculated using the distance equation of Euclide. To summarize, we synthetically report the steps followed by our algorithm (see Fig. 2):

1. *Coarse spatial filtering*: 2D image (Input no. 1) from Kinect is coarsly filtered using the depth data (Input no. 2) acquired from the IR camera to exclude undesired objects from the image. A black patch is applied to all the image except for the region containing the subject's face;
2. *FaceTracker processing*: the *FaceTracker* API identifies and localizes 66 2D landmarks on the face;


INPUT no. 1
INPUT no. 2
2D visual data
Depth data
1
Coarse spatial filtering & Use of dilated mask for 2D image improvement
2
*FaceTracker* processing
3
Landmarks selection and calculation
4
2D to 3D transformation
5
Landmark distance measurements
OUTPUT

| *No.* | *Linear measurement* |
|---|---|
| 1 | Total facial height |
| 2 | Upper facial height |
| 3 | Lower facial height |
| 4 | Intercantal width |
| 5 | Binocular width |
| 6 | Nasal height |
| 7 | Nasal projection |
| 8 | Philtrum width |
| 9 | Labial fissure width |
| 10 | Upper lip height |
| 11 | Upper lip length (left) |

**Fig. 2.** System pipeline overview.

3. *Landmark selection and calculation*: starting from the 66 localized landmarks, the 13 facial landmarks are selected and calculated;
4. *2D to 3D transformation*: using the intrinsic parameters of the Kinect camera (estimated with a preliminary offline procedure) and the depth data coming from Kinect, it is possibile to transform the 2D landmarks into the corresponding 3D points in the space;
5. *Landmark distance measurements*: the euclidean distance between a set of selected pairs of landmarks are calculated in order to extract the linear facial measurements. The output consists of 11 selected linear facial measurement expressed in millimeters.

The computational time for the entire pipeline depends on the *FaceTracker* API status. In particular, if it is the first time that the API is searching for the facial landmarks, the entire pipeline lasts 215 $ms$ (average of $10^3$ runs), while if the landmarks are already localized on the face, the pipeline lasts 24 $ms$ (average of $10^3$ runs), even if the subject moves the head[6].

[6] For pipeline computational time measurements, we use a processing system equipped with an Intel Core i7 CPU at 3.07 GHz, 6 GB RAM and Windows 7 64-bit OS.

## 3 Population and methodologies

### 3.1 Study population

The studied population is composed by 36 healthy adults, ranging from 18 to 31 years of age ($\mu = 29.97$ years, SD = 3.59 years). There are 29 men and 7 women belonging to four different nationalities. In particular, four males have beard.

### 3.2 Data collection

In this work, we considered 3 measurement methods (MM): the caliper MM, the Kinect 1 shot MM (*Kinect 1*) and the Kinect 100 shots MM (*Kinect 100*). The caliper measurements are performed by a professional operator. For every method, all 11 measurements reported in Table 1 are taken and registered. This leads to a total of 33 (3x11) facial measurements per person and, so, a total of 1188 (33x36) measurements. The first uses a digital caliper with 0.1 mm resolution and ± 0.1 mm accuracy. The second and third methods use the proposed marker-less method. In the second method, the 11 facial measurements are taken in one single shot (in 215 $ms$). On the other hand, the third method performs 100 times the same 11 facial measurements and takes, as final measurement results, the average of the measured values. The *Kinect 100* method attemps to statistically improve the precision of Kinect. Additionally, 100 shots required 2,59 seconds ($215 + 99 * 24$ $ms$), so the immobility condition of the measured person is still valid.

The entire data collection procedure for each tested person is composed by four steps. 1. Presentation and explanation of the measurement methods and collection of personal data (age, gender, values of the anthropometric facial measurements). 2. Signing the consent form for the processing of personal data. 3. Manual measurement with the digital caliper of the 11 standard facial measurements. 4. Automatic measurement of the 11 standard facial measurements using the proposed method based on Kinect (see Section 2.2) both according to *Kinect 1* and the *Kinect 100* measurement methods.

## 4 Experimental results and discussion

In Figure 3, the resulting mean and Standard Deviation (SD) of the error measurements are reported for each comparison between following method couples: *Kinect 1* and caliper methods; *Kinect 100* and caliper methods; *Kinect 1* and *Kinect 100* methods.

We perform a one-way ANOVA (ANalysis Of VAriance) test (1 degree of freedom and $\alpha = 0.05$) for each of the 11 distance measurements and for each couple of measurement methods, leading to a total of 33 combinations. These tests are performed in order to verify if there are some statistically significant differences between the measurement methods. In fact, the use of one-way ANOVA tests is admissible for our study because each facial measurement method is applied on the same population.

The results of the tests are listed in Table 2, where we synthetically report, also, if the null hypothesis can be considered valid. The validity of the null hypotesis implies that our proposed measurement methods (*Kinect 1* and/or *Kinect 100*) can be considered statistically equal to the caliper method.

Regarding the comparison between *Kinect 1* and *Kinect 100*, it is proved that they are statistically equal for each couple of facial landmarks and that the committed error in the measurement is in line with the precision of the Microsoft Kinect. It is possible to consider our system as a single-shot measurement system that after 215 *ms* provides the facial characterizing distances. In this way, the requirement of substantial immobility of the measured person is completely fulfilled. Therefore, in the following discussion, due to substantial equivalence between *Kinect 1* and *Kinect 100*, we will only compare the *Kinect 1* method to the caliper method.

According to one-way ANOVA tests, the caliper and the *Kinect 1* methods can be considered equivalent for 5 up to 11 distances among facial landmark pairs (see Table. 2): 1. ”Upper facial height”; 2. ”Nasal height”; 3. ”Labial Fissure width”; 4. ”Upper lip height”; 5. 11 ”Upper facial lenght (left)”; involving the landmarks: *n*; *sto*; *sn*; *ch* left; *ch* right. Each of these landmarks is a common point in more than one of the five aforementioned distances. Therefore, the 5 landmarks are properly identified on the face of the tested people. In addition, the subject’s beard does not interfere with the automatic landmark localization for *sto*, *sn*, *ch*s landmarks, that are placed in the mouth region.

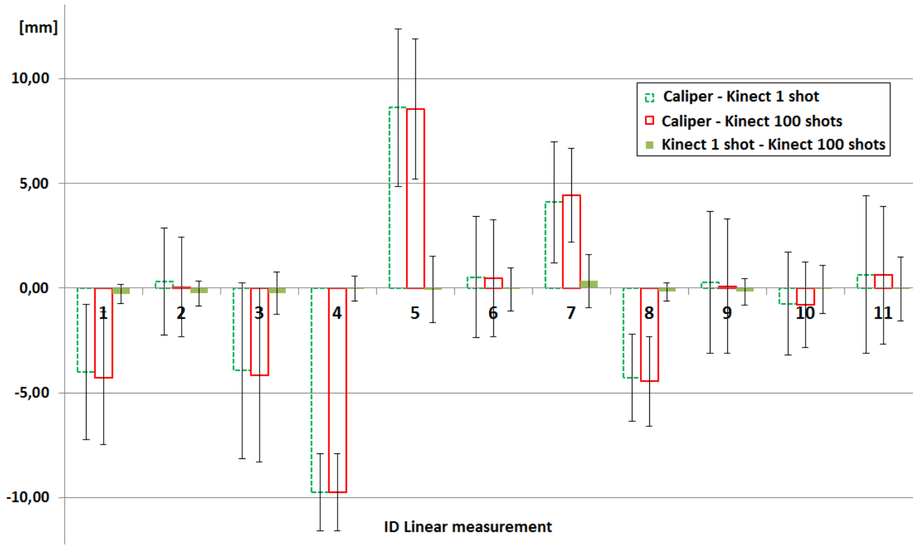


**Fig. 3.** Graph representing the mean and the SD of the measurement error resulting from the comparison between *Kinect 1* and caliper, *Kinect 100* and caliper, and *Kinect 1* and *Kinect 100* methods.

**Table 2.** Results of one-way ANOVA tests (DOFs = 1; $\alpha = 0.05$) on different couples of methods: *Kinect 1* and caliper, *Kinect 100* and caliper, and *Kinect 1* and *Kinect 100* methods.

| Measurement error | | Caliper - *Kinect 1* [mm] | | | Caliper - *Kinect 100* [mm] | | | *Kinect 1* - *Kinect 100* [mm] | | | Null hypothesis valid | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *No.* | *Landmarks* | $F$ | $\rho$ | $F_{crit}$ | $F$ | $\rho$ | $F_{crit}$ | $F$ | $\rho$ | $F_{crit}$ | Yes | No |
| 1 | n-gn | 5,272 | 0,025 | 3,978 | 6,079 | 0,016 | 3,978 | 0,029 | 0,865 | 3,978 | | x |
| 2 | n-sto | 0,063 | 0,803 | 3,978 | 0,001 | 0,977 | 3,978 | 0,052 | 0,820 | 3,978 | x | |
| 3 | sn-gn | 8,815 | 0,004 | 3,978 | 10,368 | 0,002 | 3,978 | 0,037 | 0,848 | 3,978 | | x |
| 4 | en-en | 256,504 | 0,000 | 3,978 | 258,285 | 0,000 | 3,978 | 0,000 | 0,985 | 3,978 | | x |
| 5 | ex-ex | 42,543 | 0,000 | 3,978 | 43,796 | 0,000 | 3,978 | 0,003 | 0,957 | 3,978 | | x |
| 6 | n-sn | 0,259 | 0,612 | 3,978 | 0,204 | 0,653 | 3,978 | 0,004 | 0,952 | 3,978 | x | |
| 7 | sn-prn | 46,204 | 0,000 | 3,978 | 80,559 | 0,000 | 3,978 | 0,286 | 0,594 | 3,978 | | x |
| 8 | cph-cph | 113,627 | 0,000 | 3,978 | 128,623 | 0,000 | 3,978 | 0,354 | 0,554 | 3,978 | | x |
| 9 | ch-ch | 0,121 | 0,729 | 3,978 | 0,015 | 0,904 | 3,978 | 0,046 | 0,830 | 3,978 | x | |
| 10 | sn-sto | 1,424 | 0,237 | 3,978 | 1,820 | 0,182 | 3,978 | 0,012 | 0,913 | 3,978 | x | |
| 11 | sn-ch | 0,639 | 0,427 | 3,978 | 0,689 | 0,409 | 3,978 | 0,002 | 0,965 | 3,978 | x | |

Regarding the other landmark distances that have not passed the one-way ANOVA tests (distances: "Total facial height", "Lower facial height", "Intercantal width", "Binocular width", "Nasal projection" and "Philtrum width"), we can attribute this errors to a localization displacement of the facial landmarks: *gn*, *ex*s, *en*s, *prn* and *cph*s. Indeed, these errors can not be attributed to Kinect measurements, since it has achieved good results with the distance measurements mentioned above.

The motivations for the committed errors in landmark localization are:

1. for *gn* landmark, it is necessary to palpate the chin in order to feel the bone, especially if we consider that: 1. the part between the chin and the throat is not always planar; 2. under the chin there is usually an accumulation of fat that can visually provide wrong position of the chin itself.
2. for *ex (left) and ex (right)* landmarks, the capacity of the *FaceTracker* to unproperly localize the exocanthion. In fact, the *FaceTracker* places the exocanthions over the most external points of the eyes where the eyelids touch. This leads to an underestimation of the "Binocular width" (the distance error is always positive) and to a doubled distance error (both for the left and for the right side). This diverted landmark localization can be solved through a further computer vision elaboration of the picture in the area around the exocanthion identified by the *FaceTracker* algorithm;
3. for *en (left) and en (right)* landmarks, the capacity of the *FaceTracker* to unproperly localize the endocanthion. In fact, the *FaceTracker* places the endocanthions over the eye caruncles. This leads to an overestimation of the "Intercantal width" (the distance error is always negative) and to a doubled

distance error (both for the left and for the right side). Moreover, through a further one-way ANOVA test, we prove ($F = -3E - 14$, $\rho << 0,001$ and $F_{crit} = 3,978$) that the committed error for the "Intercantal width" is systematic and equal to $9,76$ $mm$. The resulting localization displacement can be solved through a further computer vision elaboration of the picture in the area around the caruncles identified by the *FaceTracker* algorithm;

4. for *prn* landmark, the non-use of the *FaceTracker* of the 3D information for the nose. A further improvement can be to consider as correct *prn* landmark, the point belonging to a small picture area centered into the identified *prn*, but nearest to Kinect (considering the facial plane perpendicular to the Kinect);
5. for *cph (left) and cph (right)* landmarks, the numerosity of different shapes and colors of lips and the colors of the skin around the lips. Even in the measurements taken with the caliper, the cph landmarks were most of the times difficult to localize, especially when: 1. the *cph*s did not have a pointed shape; 2. the upper part of the lip and the skin had the same color. Additionally, through a further one-way ANOVA test, we rejected the hypothesis according to the error is due to beard.

In order to improve the proposed method, an enhancement of the *FaceTracker* API is required. It can be done applying further computer vision algorithms (such as those involving Laplacian and gradient filters) and using 3D information from the Kinect. This can lead to solve the localization landmark displacements.

Finally, only by using the caliper measurements, we have statistically rejected the hypotesis reported in many works of the literature that states the equivalence of the following equations: $dist_{n-gn} = dist_{n-sn} + dist_{sn-gn} \Leftrightarrow Totalfacialheight = Nasalheight + Lowerfacialheight$.

Measurement error statistics ($\mu$ = -4.28, $SD$ = 3.58, $F$ = 5,226, $\rho$ = 0,025 and $F_{crit}$ = 3,978) show that the group *n-gn* and the group (*n-sn* + *sn-gn*) are significantly different.

## 5 Conclusion and future works

In this paper we introduced and discussed a new marker-less 3D Kinect-based system for facial anthropometric measurements. It allows to overcome the main drawbacks which characterize the existing facial measurement solutions. Commonly used solutions require an high professional human intervention to correctly localize the facial landmarks. Moreover, in some cases, it is expected to fix the markers on the subject's face, leading to an increase of the degree of intrusiveness of the measurement system.

Furthermore, our adopted approach allows to keep the cost of the system low and to perform the measurement process in 215 $ms$, avoiding any potential error due to head movements. The experimental part revealed a successful percentage of our method for 54,5 % of the total measured distances with respect to the caliper-based manual system (considering the systematic error committed on "Intercantal width"). The increase of the performances, as well as the decrease of magnitude and variance of the measurement errors are the aim of future works. Considering the causes of the landmark localization displacements presented in Section 4, we want to improve the automatic localization of the facial landmarks and speed up the pipeline process with a GPU version of the algorithm.

**Acknowledgments.** This work is partially supported by Instituto de Telecomunicações, Fundação para a Ciência e Tecnologia (SFRH/BD/69878/2010, SFRH/BD/33974/2009), VERE project (ref: 257695), LIFEisGAME project (ref: UTA-Est/MAI/0009/2009) and Golem project (ref.251415, FP7-PEOPLE-2009-IAPP). Special Thanks go to Jacqueline Fernandes for helping with the editing of the article.

## References

1. Sinnatamby C.: Last's anatomy. Regional and applied, vol. 10 (1999)
2. Choe K., Sclafani A., Litner J., Yu G., Romo T.: The korean american woman's face. Archives of facial plastic surgery vol 6, no. 4, p. 244 (2004)
3. Bianchini E. M. G., Avaliaçao fonoaudiológica da motricidade oral-disturbios miofuncionais orafaciais ou situaçöes adaptativas; Speech-pathologist evaluation-orofacial myofunctional disorders or compensatory situation. Rev. dent. press ortodon. ortop. maxilar, vol. 6, no. 3, pp. 73–82 (2001)
4. Ward R., Jamison P., Allanson J.: Quantitative approach to identifying abnormal variation in the human face exemplified by a study of 278 individuals with five craniofacial syndromes. American journal of medical genetics, vol. 91, no. 1, pp. 8–17 (2000)
5. Porter J., Olson K.: Anthropometric facial analysis of the african american woman. Archives of Facial Plastic Surgery, vol. 3, no. 3, p. 191 (2001)
6. DeCarlo D., Metaxas D., Stone M.: An anthropometric face model using variational techniques. In: Proceedings of the 25th annual conference on Computer graphics and interactive techniques. ACM, pp. 67–74 (1998)
7. Kau C., Richmond S., Zhurov A., Knox J., Chestnutt I., Hartles F., Playle R.: Reliability of measuring facial morphology with a 3-dimensional laser scanning system. American journal of orthodontics and dentofacial orthopedics, vol. 128, no. 4, pp. 424–430 (2005)
8. Hammond P., Hutton T., Allanson J., Campbell L., Hennekam R., Holden S., Patton M., Shaw A., Temple I., Trotter M.: 3d analysis of facial morphology. American Journal of Medical Genetics Part A, vol. 126, no. 4, pp. 339–348 (2004)
9. Schimmel M., Christou P., Houstis O., Herrmann F., Kiliaridis S., Müller F.: Distances between facial landmarks can be measured accurately with a new digital 3-dimensional video system. American Journal of Orthodontics and Dentofacial Orthopedics, vol. 137, no. 5, pp. 580–e1 (2010)
10. Weinberg S., Scott N., Neiswanger K., Brandon C., Marazita M.: Digital three-dimensional photogrammetry: evaluation of anthropometric precision and accuracy using a genex 3d camera system. The Cleft palate-craniofacial journal, vol. 41, no. 5, pp. 507–518 (2004)
11. Saragih J., Lucey S., Cohn J.: Deformable model fitting by regularized landmark mean-shift. International Journal of Computer Vision, pp. 1–16 (2011)