Pattern-based Face Localization and Online Projector Parameterization for Multi-Camera 3D Scanning

Karima Ouji*[a], Mohsen Ardabilian[a], Liming Chen[a] and Faouzi Ghorbel[b]
[a]LIRIS Laboratory, Ecole Centrale de Lyon, France
[b]GRIFT Laboratory, Ecole Nationale des Sciences de l'Informatique, Tunisie

Abstract

3D face modeling is a very important and challenging task in computer vision especially in the presence of an expression variation. It plays an important role in a wide range of applications including facial animation, human-computer interaction and surgery. Current 3D face imaging solutions are often based on structured light camera/projector systems to overcome the relatively uniform appearance of skin. This paper proposes a space-time scheme for 3D face scanning, employing three calibrated cameras coupled with a non calibrated projector device. The proposed solution is a hybrid stereovision and phase-shifting approach, using two π-shifted sinusoid patterns and a texture image. It involves a pattern-based face localization approach and an online projector parameterization. The experimental results further validated the effectiveness of the proposed approach.

Keywords: Stereovision, Phase-shifting, Object Localization, Projector parameterization, 3D scanning.

1. Introduction

3D face modeling is a very important and challenging task in computer vision especially in the presence of an expression variation. It plays an important role in a wide range of applications including facial animation, human-computer interaction and surgery [1]. Current 3D face imaging solutions are often based on structured light camera/projector systems to overcome the relatively uniform appearance of skin [2-4]. Depth information is recovered by decoding patterns of a projected structured light which include gray codes, sinusoidal fringes, etc. Current solutions mostly utilize more than three phase-shifted sinusoidal patterns to recover the depth information, thus impacting the acquisition delay; they further require projector-camera calibration whose accuracy is crucial for phase to depth estimation step; and finally, they also need an unwrapping stage which is sensitive to ambient light, especially when the number of patterns decreases [5]. An alternative to projector-camera systems consists of recovering depth information by stereovision using a multi-camera system as proposed in [4, 6]. Here, a stereo matching step finds correspondence between stereo images and the 3D information is obtained by optical triangulation [4, 7]. Meanwhile, the model computed in this way generally is quite sparse. Recently, researchers looked into super-resolution techniques as a solution to upsample and denoise depth images. Kil et al. [8] applied super-resolution for laser triangulation scanners by regular resampling from aligned scan points with associated Gaussian location uncertainty. Super-resolution was especially proposed for time-of-flight cameras which have very low data quality and a very high random noise by solving an energy minimization problem [9].

This paper proposes a space-time scheme for 3D face scanning, employing three calibrated cameras coupled with a non calibrated projection device. The proposed solution is a hybrid stereovision and phase-shifting approach, using two π-shifted sinusoid patterns and a texture image. It involves a pattern-based face localization approach to decrease the whole processing time. This work suggests as well an online projector parameterization and does not require a camera-projector off-line calibration which constitutes a tedious and expensive task. In contrast to conventional structured-light methods, our method avoids the phase unwrapping stage thanks to the use of stereo in the first stage of the approach. Moreover, a temporal super-resolution is proposed to correct the 3D information, to complete the 3D scanned view and to consider the facial deformable aspect. Section (2) presents the system overview. Section (3) details the pattern-based face localization. In section (4), we high-light the online projector parameterization. Section (5) explains how the 3D space-time super-resolution is carried out. Section (6) discusses the experimental results and section (7) concludes the paper.

* karima.ouji@ec-lyon.fr; http://liris.cnrs.fr

2. System Overview

Figure 1 presents our 3D face scanning scheme. First, an offline strong stereo calibration computes the intrinsic and extrinsic parameters of the cameras, estimates the tangential and radial distortion parameters, and provides the epipolar geometry as proposed in [10]. In online process, two π-shifted sinusoid patterns and a third white pattern are projected onto the face. Three sets of left, middle and right images are captured, undistorted and rectified. The proposed model is defined by the system of equations (1) and allows the computation of both the sparse and dense models. It constitutes a variant of the mathematic model proposed in [5].

$$\begin{aligned} I_p(s,t) &= I_b(s,t) + I_a(s,t)\cdot\sin(\phi(s,t)), \\ I_n(s,t) &= I_b(s,t) + I_a(s,t)\cdot\sin(\phi(s,t)+\pi), \\ I_t(s,t) &= I_b(s,t) + I_a(s,t). \end{aligned} \tag{1}$$

At time t, $I_p(s,t)$, $I_n(s,t)$, $I_t(s,t)$ constitute the intensity value of the pixel s on respectively the positive image, the negative one and the texture one. $I_b(s,t)$ represents the texture information and the lighting effect. $\phi(s,t)$ is the local phase defined at each pixel s. Solving (1), $I_b(s,t)$ is computed as the average intensity of $I_p(s,t)$ and $I_n(s,t)$. $I_a(s,t)$ is then computed from the third equation of the system (1) and ϕ(s,t) is estimated by equation (2).

$$\phi(s,t) = \arcsin\left[\frac{I_p(s,t)-I_n(s,t)}{2\cdot I_t(s,t)-I_p(s,t)-I_n(s,t)}\right]. \tag{2}$$

A 3D sparse model is estimated from stereo-matching with a fringe-based resolution and a sub-pixel precision [7]. Second, a precise projector parameterization is performed based on the computed sparse model. Thus, we avoid the camera-projector off-line calibration which constitutes a tedious and expensive task.

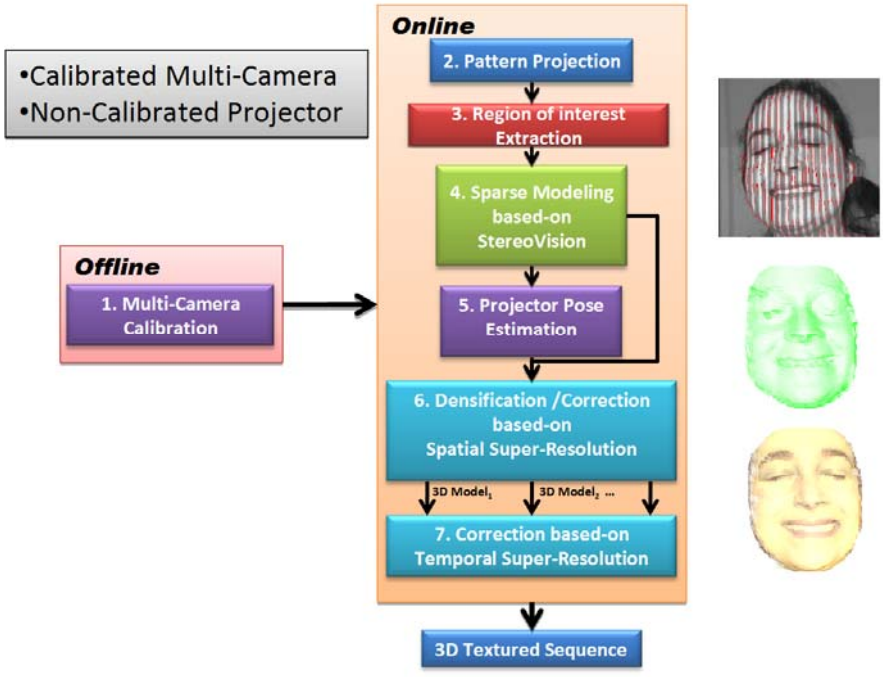

Fig.1. 3D Sequence Acquisition Framework of a textured moving face

Then, a dense 3D model is recovered by the intra-fringe phase estimation, from the two sinusoidal fringe images and a texture image, independently from the left, middle and right cameras. The left, middle and right 3D dense models are fused to produce the final 3D model which constitutes a spatial super-resolution. Moreover, a temporal super-resolution based on the CPD algorithm is proposed to

correct the 3D information, to complete the 3D scanned view and to consider the facial deformable aspect [4]. In contrast to conventional methods, our method avoids the phase unwrapping stage thanks to the use of stereo in the first stage of the approach. The experimental results further validated the effectiveness of the proposed approach.

3. Pattern-based face localization

When one sinusoidal pattern is projected on the face, a strong contrast occurs on the informative facial area and weak contrast characterizes the background of the scanned face. The idea is to benefit from the contrast variation to localize the region of interest and carry out a spectral analysis to localize the low frequencies on captured images. Figure 2.a presents a captured image and figure 2.c presents the face localization result. To localize the facial region, we compute FFT on a sliding window for each epiline which provides for each pixel a 2D curve of FFT frequency amplitudes. A 3D spectral distribution is obtained which highlights the facial region for the current epiline as shown in figure 2.b. We propose to keep only pixels belonging to this highlighted region. Thus, for each pixel in the epiline, we consider a weighted sum of only the low-frequency amplitudes and we apply an adequate thresholding to obtain the region-of-interest as illustrated by figure 2.d.

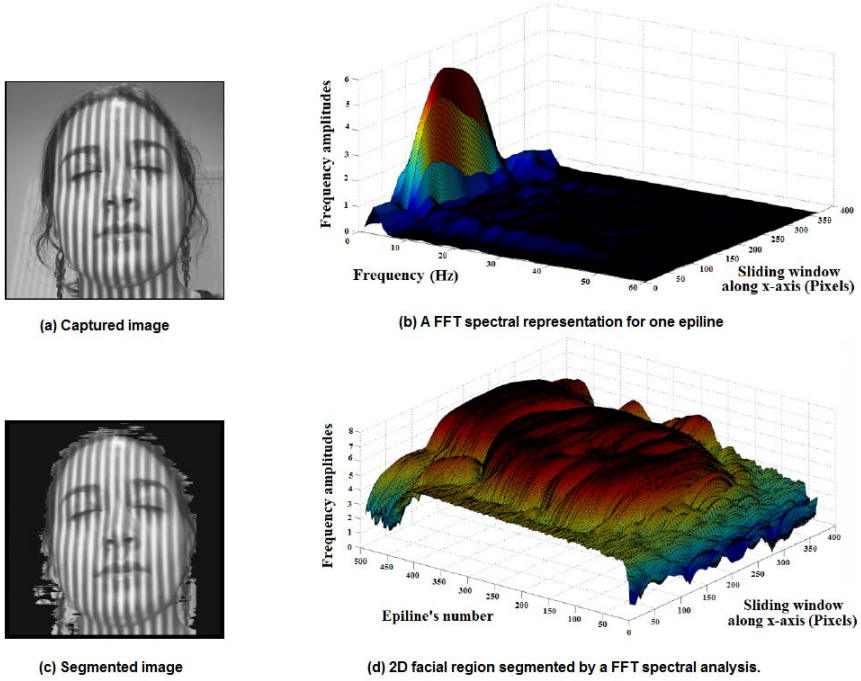

Fig.2. Pattern-based face localization

4. 3D sparse modeling for online projector parameterization

When projecting vertical fringes, the video projector can be considered as vertical adjacent sources of light. Such a consideration provides for each epiline a light source point O_{Prj} situated on the corresponding epipolar plane. Our projector parameterization is performed based on the computed sparse model generated through a stereovision scenario. The sparse model is formed by the primitives situated on the fringe change-over which is the intersection of the sinusoidal component of the positive image and the second π-shifted sinusoidal component of the negative one [6]. Therefore, the localization has a sub-pixel precision. Corresponding multi-camera primitives necessarily have the same Y-coordinate in the rectified images. Thus, stereo matching problem is resolved in each epiline separately using Dynamic Programming. The 3D sparse point cloud is then recovered by computing the intersection of optical rays coming from the pair of matched features.

The sparse 3D point cloud is a serie of adjacent 3D vertical curves obtained by the fringes intersection of the positive and the negative images. Each curve describes the profile of a projected vertical fringe distorted on the 3D facial surface. We propose to estimate the 3D plane containing each distorted 3D curve separately. As a result, the light source vertical axis of the projector is defined as the intersection of all the computed 3D planes. This estimation can be performed either as an offline or online process

unlike conventional phase-shifting approaches where the projector is calibrated on offline and cannot change its position when scanning the object.

5. Non-rigid depth super-resolution

Here the objective is to densify the sparse model obtained from active stereo-vision. We assume that the vertical axis of the projection source is accurately estimated. Our approach consists in: first, to estimate separately for each camera the depth information of the intrafringes pixels from the observed phase information and second, to fuse them to achieve a more complete and precise model. Concerning the first step, the 3D coordinates of each pixel are computed using phase-shifting analysis. Conventional phase shifting techniques estimate the local phase in [0..2π] for each pixel on the captured image; the absolute phases are calculated by phase unwrapping of theses local phases, also defined as wrapped phases. Finally the depth information is recovered through a phase to depth conversion which is based on the geometry of the camera-projector pair as described in [11, 12]. In the proposed approach, the sparse model lets us retrieve 3D intra-fringe information from wrapped phases directly using an adequate phase-to-depth algorithm that we proposed in [14]. Conventional super-resolution techniques carry out a registration step between low-resolution data, a fusion step and a deblurring step. Here, the phase-shifting analysis provides aligned left, middle and right point clouds since their 3D coordinates are computed based on the same 3D sparse point cloud. Also, left, middle and right point clouds present homogeneous 3D data and need only to be merged to retrieve the high-resolution 3D point cloud.

A 3D face model can present some artifacts caused by an expression variation, an occlusion or even a facial surface reflectance. To deal with these problems, we propose to apply a 3D temporal super-resolution for each couple of successive 3D point sets M_{t-1} and M_t at each moment t. First, a 3D non-rigid registration is performed and formulated as a maximum-likelihood estimation problem since the deformation between two successive 3D faces is non rigid in general. We employ the CPD (Coherent Point Drift) algorithm proposed in [11] to registrate the 3D point set M_{t-1} with the 3D point set M_t. The CPD algorithm considers the alignment of two point sets M_{src} and M_{dst} as a probability density estimation problem and fits the GMM (Gaussian Mixture Model) centroids representing M_{src} to the data points of M_{dst} by maximizing the likelihood as described in [11]. The core of the CPD method is forcing GMM centroids to move coherently as a group, which preserves the topological structure of the point sets. The coherence constraint is imposed by explicit reparameterization of GMM centroid locations for rigid and affine transformations. For smooth non-rigid transformations like expression variation, the algorithm imposes the coherence constraint by regularization of the displacement field [12].

We propose to test the efficiency of the CPD non-rigid algorithm especially if the point sets contain a high random noise and insert some ICP (Iterative Closest Point Algorithm) iterations to minimize false points and facilitate the convergence. Once registered, the 3D point sets M_{t-1} and M_t and also their corresponding 2D texture images are used as a low resolution data to create a high resolution 3D point set and its corresponding texture. We apply the 2D super-resolution technique as proposed in [13]. The 3D model M_t cannot be represented by only one 2D disparity image since the points situated on the fringe change-over have sub-pixel precision. Also, the left, middle and right pixels participate separately in the 3D model since the 3D coordinates of each pixel are retrieved using only its phase information. Thus, we propose to create for each camera three 2D maps defined by the X, Y and Z coordinates of the 3D points. The optimization algorithm and the deblurring are applied for each camera separately to compute high-resolution images of X, Y, Z and texture from the low-resolution images. We obtain for each camera a high-resolution 3D point cloud using high-resolution data of X, Y and Z. The final high-resolution 3D point cloud is retrieved by merging the left, middle, and right obtained 3D models which are already registrated since all of them contain the 3D sparse point cloud.

6. Experimental results

The stereo system hardware is formed by three network cameras with 30 fps and a 480x640 pixel resolution and a LCD video projector with a 1024x768 pixel resolution. Scanning one 3D frame needs three successive images with a different projected pattern. For a moving face, the obtained three successive images present a slight expression variation able to create some artifacts or a depth imprecision. The cameras and the projector are not synchronized by a special hardware which leads to a bad pattern projection on the face. The projector vertical axis is defined by a directional 3D vector

$\overrightarrow{N_{proj}}$ and a 3D point P_{proj}. $\overrightarrow{N_{proj}}$ and P_{proj} are computed by analyzing each 3D plane defined by all the 3D points situated at the vertical profile of the same fringe change-over. The equations of the fringe planes are estimated by a mean square optimization method with a precision of 0.002mm. The directional vector $\overrightarrow{N_{proj}}$ is then computed as the normal vector to all the normal vectors of the fringe planes. $\overrightarrow{N_{proj}}$ is estimated with a deviation error of 0.003rad. Finally, the point P_{proj} is computed as the intersection of all the fringe planes using a mean square optimization.

Figure 3 presents the primitives extracted and the reconstruction steps to create one facial 3D view with neutral expression from only two cameras. The precision of the reconstruction is estimated using a laser 3D face model scanned by a MINOLTA VI-300 non-contact 3D digitizer. We perform a point-to-surface variant of the 3D rigid matching algorithm ICP (Iterative Closest Point) between a 3D face model provided by our approach and a laser 3D model of the same face. The mean deviation obtained between them is 0.3146mm.

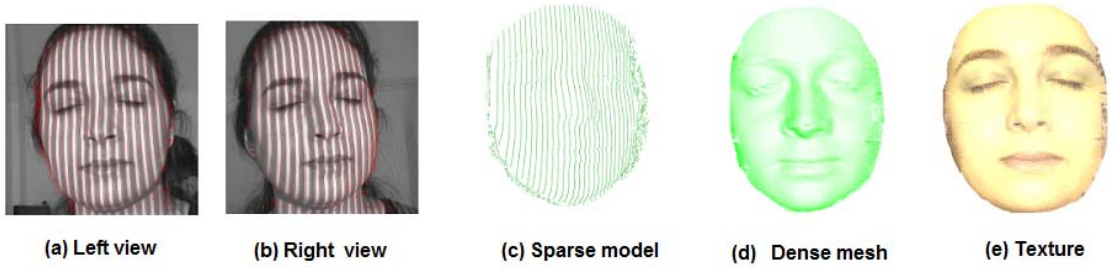

Fig.3. Reconstruction steps to create one facial 3D view from only two cameras.

The proposed approach needs a non-rigid matching step between the 3D frame F_t and its preceding 3D frame F_{t-1}. The CPD performs an efficient matching in the presence of non-rigid deformations. Figure 4 shows the result of a CPD rigid matching between two successive 3D frames with an expression variation and the result of a CPD non-rigid matching between the same faces. It describes the spatial deviation after the matching process and illustrates the efficiency of the CPD non-rigid algorithm. The non-rigid matching provides a mean spatial deviation of 0.0387mm/pixel and a standard deviation of 0.0371mm/pixel. The rigid matching provides a mean spatial deviation of 0.0616mm/pixel and a standard deviation of 0.0566mm/pixel.

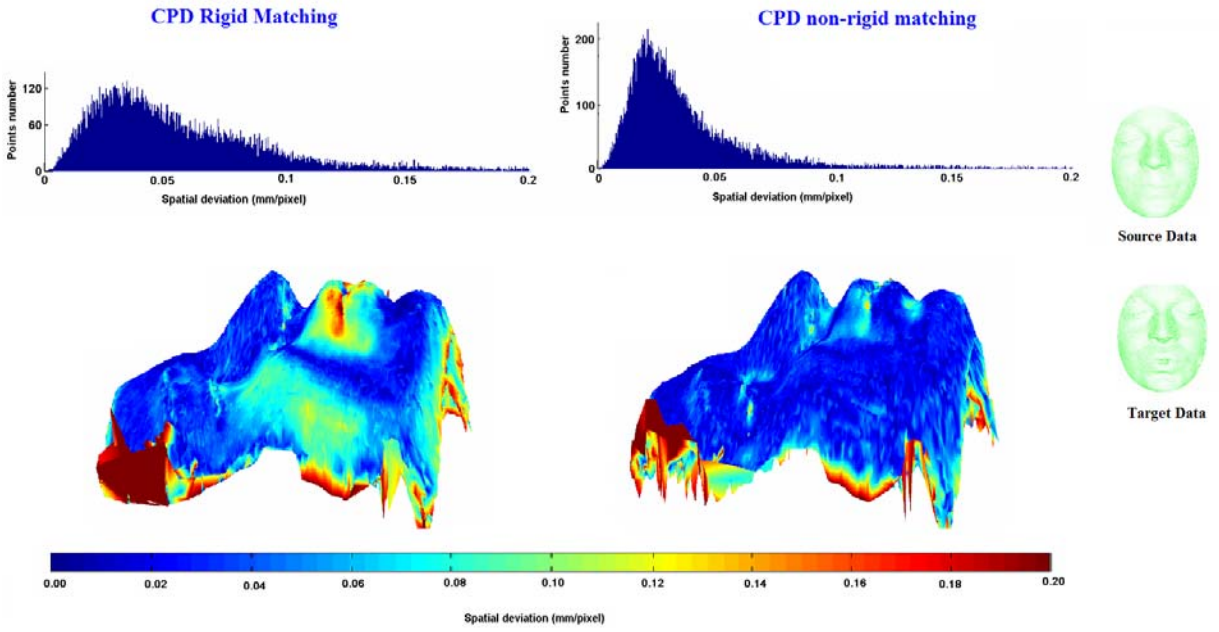

Fig.4. CPD Rigid and non-rigid matching results

At time t, the left, middle and right cameras capture three different 2D views as shown in figure 5, each view is represented by a set of three images containing respectively the positive pattern, the π-shifted pattern and the white pattern. These 2D views provide two 3D facial views which can present some artifacts as shown in figure 6 especially for the left 3D view of the second 3D frame shown in 6.d. To deal with these errors, the first and second 3D frames are merged despite their non-rigid deformation thanks to the super-resolution approach proposed in section (5).

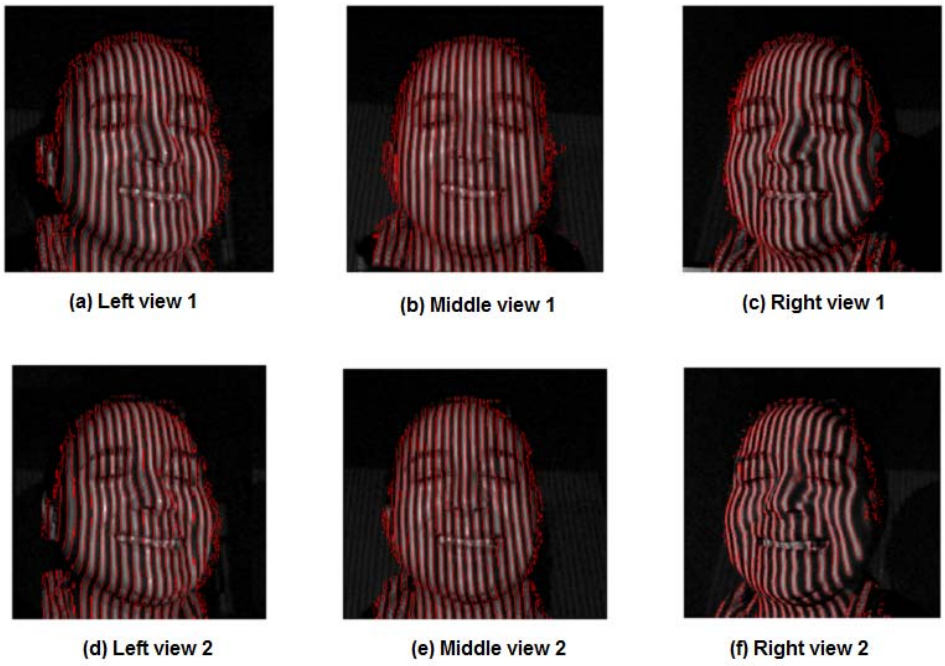

Figure 5: Sampled primitives on left, middle and right views for two successive frames with an expression variation

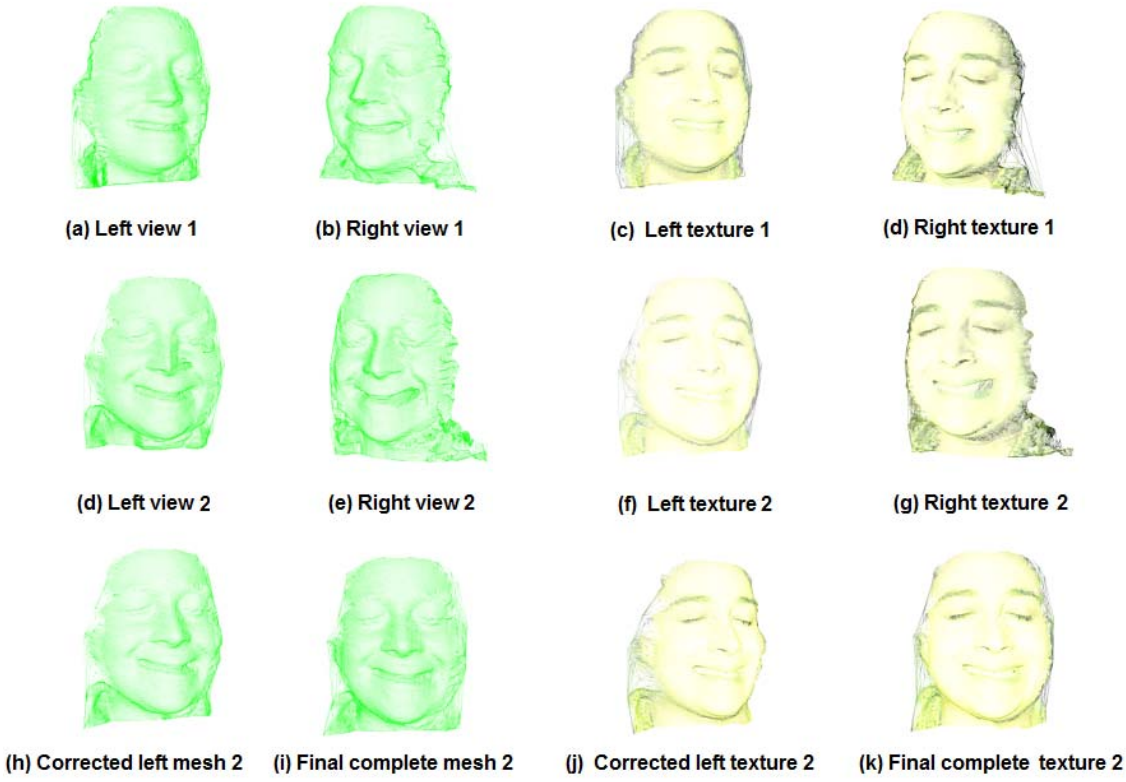

Figure 6: 3D space-time results.

As shown in figure 7, the non rigid matching algorithm CPD matches the preceding 3D frame with the current 3D left view with a mean deviation of 0.1222mm/pixel. When some ICP iterations are employed with the CPD non-matching iterations, an efficient matching result is obtained with a mean deviation of 0.0437mm/pixel. Also, the hybrid (ICP+CPD) algorithm localizes and clears the artifacts which represent a high spatial deviation with the preceding 3D frame.

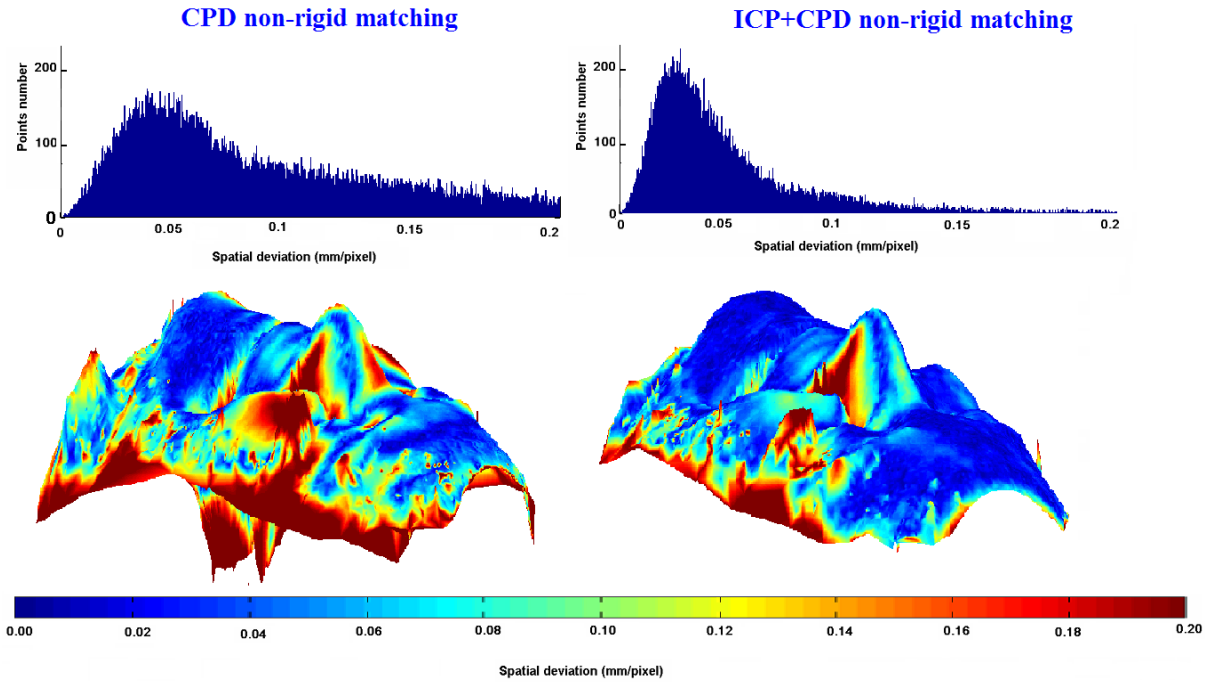

Fig.7. Non-rigid matching results

7. Conclusion and future work

This paper proposes a multi-camera 3D acquisition solution with a 3D space-time super-resolution scheme which is particularly suited to 3D face scanning. It involves a pattern-based face localization approach to decrease the whole processing time. This work suggests as well an online projector parameterization and does not require a camera-projector off-line calibration which constitutes a tedious and expensive task. We develop an efficient non-rigid registration approach to deal with the facial deformable behavior especially in the presence of an expression variation. As a future work, we suggest considering more 3D frames through the time axis. Also, hardware synchronization between the cameras and the projector will be performed to get an efficient and real-time 3D video scanning result.

References

1. Huang, D., Ouji, K., M. Ardabilian, Wang, Y. And Chen, L., (2011): "3D Face Recognition based on Local Shape Patterns and Sparse Representation Classifier", Int. Conf. on MultiMedia Modeling.
2. Zhang, S., (2010): "Recent progresses on real-time 3D shape measurement using digital fringe projection techniques", J. Optics and Lasers in Engineering, vol 48, pp 149-158.
3. Zhang, S. and Huang, P.S., (2006): "High-resolution, real-time three-dimensional shape measurement", J. Optical Engineering. 45 (123601).
4. Zhang, L., Curless, B and Seitz, S. M., (2002): "Rapid shape acquisition using color structured light and multipass dynamic programming", 3DPVT Conference.
5. Zhang, S. and Yau, S., (2008): "Absolute phase-assisted three-dimensional data registration for a dual-camera structured light system", J. Applied Optics, 47:3134–3142.
6. Cox, I., Hingorani, S. and Rao, S., (1996): "A maximum likelihood stereo algorithm". J. Computer Vision and Image Understanding, 63:542–567.
7. Ouji, K., Ardabilian, M., Chen, L. and Ghorbel, F., (2009): "Pattern Analysis for an Automatic and Low-Cost 3D Face Acquisition Technique", IEEE Advanced Concepts for Intelligent Vision Systems Conference (ACIVS), Bordeaux, France.

8. Kil, Y., Mederos, Y. and Amenta N., (2006): “Laser scanner super-resolution”, Eurographics Symposium on Point-Based Graphics.
9. Schuon, S., Theobalt, C., Davis, J., and Thrun, S., (2009): “Lidarboost: Depth superresolution for TOF 3D shape scanning”, CVPR Conference.
10. Zhang, Z., (1999) : “Flexible camera calibration by viewing a plane from unknown orientations”, ICCV Conference.
11. Myronenko, A., Song, X. and Carreira-Perpinan, M. A., (2007): “Non-rigid point set registration: Coherent point drift”, NIPS Conference.
12. Myronenko, A. and Song, X., (2010): “Point set registration: Coherent point drift”, IEEE Trans. PAMI, vol 32, pp 2262–2275.
13. Farsiu, S., Robinson, D., Elad, M. and Milanfar, P., (2004): “Fast and robust multi-frame superresolution”, IEEE Trans. Image Processing.
14. Ouji, K., Ardabilian, M., Chen, L. and Ghorbel, F., (2011): “A Space-Time Depth Super-Resolution Scheme For 3D Face Scanning”, IEEE Advanced Concepts for Intelligent Vision Systems Conference (ACIVS), Het Pand, Ghent, Belgium.
15. Ouji, K., Ardabilian, M., Chen, L. and Ghorbel, F., (2011): “Multi-Camera 3D Scanning with a Non-rigid and Space-Time Depth Super-Resolution capability”, IAPR International Conference on Computer Analysis of Images and Patterns (CAIP), Seville, Spain.