SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
AIVP ADVANCES IN IMAGE AND VIDEO PROCESSING
VOLUME 3, ISSUE 5
ISSN 2054-7412

Patches Detection and Fusion for 3D Face Cloning

Jérôme Manceau, Catherine Soladié and Renaud Séguier
Supélec, Team FAST (Facial Analysis, synthesis and tracking), Rennes, France;
Avenue de la Boulaie, 35510 Cesson-Sévigné
Jerome.manceau@supelec.fr; catherine.soladie@supelec.fr; renaud.seguier@supelec.fr

ABSTRACT

3D face clones are used in many fields such as video games and Human-Computer Interaction. However, high-resolution sensors generating high quality clones are expensive and not accessible to all. In this paper, we propose to make a fully automated and accurate 3D reconstruction of a face with a low cost RGB-D camera. For each subject, we capture the depth and RGB data of their face in different positions while performing a rotational movement of the head. We fit a 3D Morphable Face Model on each frame to eliminate noise, increase resolution and provide a structured mesh. This type of mesh is a mesh which the semantic and topological structure is known. We propose to only keep the suitable parts of each mesh called Patch. This selection is performed using an error distance and the direction of the normal vectors. To create the 3D face clone, we merge the different patches of each mesh. These patches contain relevant information on the specificity of individuals and lead to the construction of a more accurate clone. We perform quantitative tests by comparing our clone to ground truth and qualitative tests by comparing visual features. These results show that our method outperforms the FaceWarehouse process of Cao et al [2]. This 3D face clone on a structured mesh can be used as pretreatment in applications such as emotion analysis [13] or facial animation.

Keywords: Structured mesh, Patches detection and fusion, 3D Morphable Face Model, Fitting

1 Introduction

Face Cloning is an important area of research in Computer Vision and Graphics. Indeed, it can be used in many applications, such as video and serious games, e-learning and Human Computer Interaction where the user must be able to interact with the computers. Actually, these applications must assist machines to automatically detect specific information about the user such as hand, arm and face gestures. A lot of research is conducted to improve such applications. R.Gross et al [8], C.Soladie et al [11] and A.Väljamäe et al [12] show that systems which adapt to the specificities of the subjects perform better than generic systems. For this reason, the use of a 3D face clone of the user rather than a generic face model as pretreatment increases the performance of these applications, K.A.Funes Mora and J.Odobez[7] shows for example that the use of a 3D clone to detect the pose of the head and eyes provides excellent results. That is why to improve the reconstruction techniques; realistic 3D clone can increase the performance of these applications. Moreover, the necessary infrastructure shall be available to end-users at their homes. Therefore, the sensor must be inexpensive and the method must be fully automatic. For all these reasons, low-resolution cameras have been recently used in the field of facial clones. Furthermore, we must know the semantic and topological structure of meshes so that the 3D clones can be used in applications. We call this type of mesh, a structured mesh. For instance structured mesh allows the detection of characteristic points used identifying a person's emotions [13].

DOI: 10.14738/aivp.35.814
Publication Date: 14th November, 2015
URL: http://dx.doi.org/10.14738/aivp.35.814

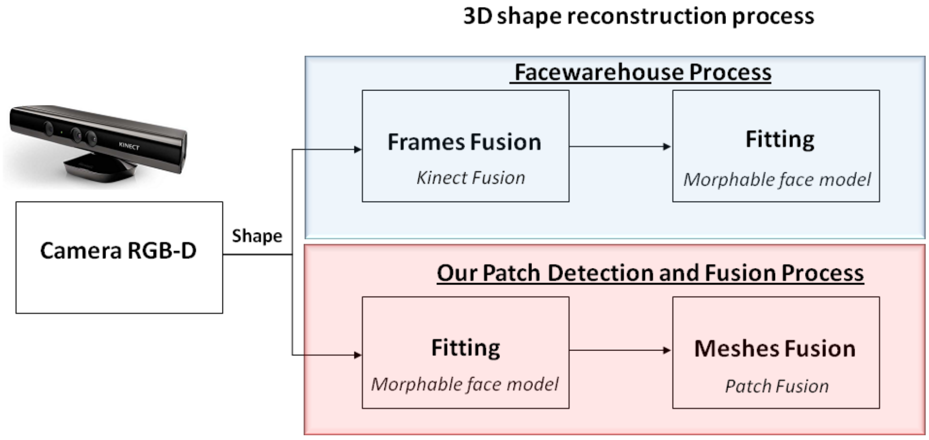

Figure 1: Comparison of 3D shape reconstruction process

There are several types of sensors for obtaining realistic clones. In the literature, certain methods use 2D data (RGB) to reconstruct the 3D shape and texture. Light Stage is a 2D high-resolution scanner that captures the properties of light (in texture and reflectance) of any object. This technology was developed in California by P.Debevec [5]. This method uses the specific properties of the skin. It consists of several light sources (LED), several digital cameras and electronic system for controlling the light and the RGB camera. Highly realistic clones can be obtained but is very expensive and not accessible. The web service AutoDesk 123D Catch permits to create realistic clones from 2D images but it does not provide structured clones. In addition, we need the help of a second person to create his clone. There are also several types of high-resolution 3D sensors for obtaining hyper realistic clones. This type of sensor is used to achieve very satisfactory results in terms of accuracy and realism but they are not feasible for domestic use. They are used to create databases and ground truth. Indeed Inspeck Mega Capturor II 3D has created the basis of the Bosphorus data with an accuracy of 0.3 mm ref [9]. It makes it possible to acquire depth data with structured light. P.Paysan et al [10] use a coded light system created by ABW-3D. They measure the shape of an object using a sequence of light patterns. This scanner provides realistic clones with high resolution. It was used to design the database of their Morphable Face Model [10].

In this paper, we propose a system for 3D face cloning using a low-cost sensor (Kinect) and providing a structured high resolution 3D clone. With this sensor, we obtain noisy low-resolution depth and color data. Therefore, we fit a Morphable Face Model [10] on each 3D depth frame (Figure 2). This has two advantages: 1) it enables to increase the resolution and reduce the noise for each 3D depth frame 2) it enables to know the structure of 3D facial mesh. We obtain for each frame a structured 3D mesh. Our process is completely automatic: we have no manual training phase. In response to realism, we propose a method for detecting and merging parts of the obtained meshes (patches) that are adequate. Indeed, we identify patches contain relevant information on the specificity of an individual. Finally, we merge all of these patches. This approach allows us to provide a realistic 3D clone.

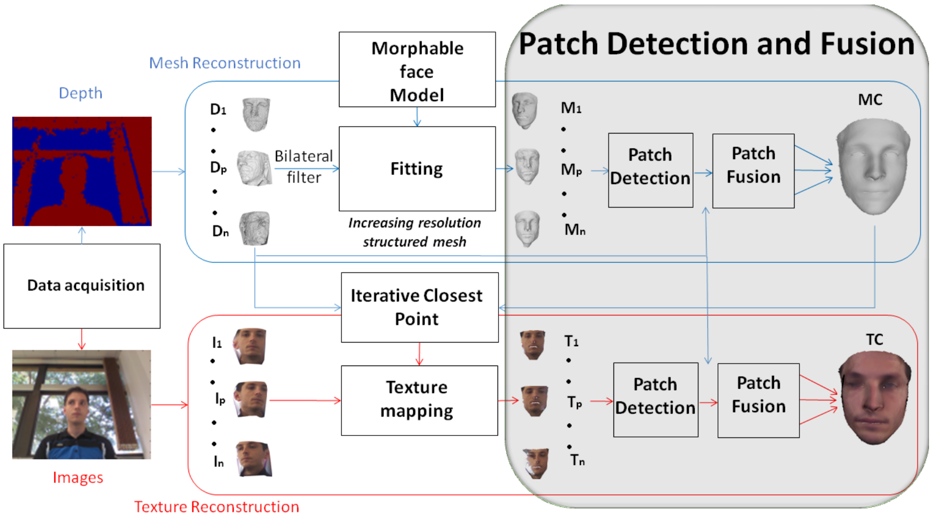

Figure 2: Our patch fusion for 3D face cloning system

The first main contribution of this paper is at the system level. Most of the methods first perform frames fusion and then a Morphable Face Model (Figure 1). In the FaceWarehouse process, C. Cao et al [2] fit a Morphable Face Model on a mesh obtained with Kinect Fusion. The peculiarity of our method is to reverse the process. We first perform the fitting on each depth frame and then the fusion. Under these conditions, the system is less dependent on alignments and fitting errors because we merge a posteriori only reliable information. Our second main contribution is the patches detection and fusion technique. When we use a Morphable Face Model, some of the morphological specificities of the individual that we want to clone may disappear. Indeed, the entire Morphable Model does not contain all possible forms and details of the unknown new face in its entirety. The specificities of the individual can only be found if they belong to the database. That is why we used a method that selects small patches (carefully chosen set of points) that focus on the details of the specificities of the individual. That is, we identify the parts (patches) of each 3D meshes that are relevant using an error distance and the direction of the normal vectors at each point of the face. Our approach allows finding specificities of persons that are not found with a conventional method of fitting [16]. We use this method both on the texture and on the depth data.

This article is organized as follows. In the next section, we present several methods for cloning 3D face that exist in the literature. In Part 3, we describe the various components of our patches detection and fusion algorithm on form and texture. Part 4 demonstrates the accuracy and the precision of our results by comparing them to other methods. Section 5 concludes the paper.

2 Related Work

There are several techniques for cloning 3D face with low resolution RGB-D sensors. During the past decade, these sensors have often been used in research because they aren't expensive and are accessible to general public. These methods can be classified into two categories of methods: techniques to obtain an unstructured mesh and those which give structured one. A structured mesh is a mesh for which we know the correspondence of each 3D point with the face that we want to clone (figure 3).

R.A Newcombe et al [17] present a 3-D reconstruction of scenes or objects using a depth low cost camera. Kinect Fusion provides high quality 3D scans. The algorithm consists of 3 steps. First Iterative Closest Point algorithm is used to determine the position of the camera. Then, they use a surface volumetric representation [18]. And finally they perform a ray casting for rendering depth data. To increase the resolution of the depth map, Y. Cui et al [4] use a method of super-resolution [19]. This method creates a high resolution depth map from multiple low resolution depth maps. It combines low resolution depth maps with a perspective slightly different from the static object. This approach gives less noisy and smoothed frames. Then each frame is aligned to reconstruct the 3D face using a probabilistic alignment method [20]. Q. Sun et al [15] propose a method for reconstructing a 3D face from RGB and depth data captured with low resolution Kinect. First, it detects the person's face by using the RGB data. Then they use bilinear interpolation to increase the resolution of depth frames. Finally, they combine four frames high resolution depth for a realistic 3D face. They use an energy function that will allow to combine the frames of depth but also to smooth the final result. All these methods enable to clone realistic faces but are limited in terms of precision of facial features. For example, the low resolution sensors do not reflect the accurate shape of the face at the vicinity eyes. Because of the infrared reflection in the eyes, the sensor does not return the shape of the eyeball (Figure 11). Moreover, they do not provide structured 3D clones and therefore cannot be directly used as a pretreatment in applications requiring knowledge of the correspondence of the mesh points with the face. Below, we present the teams that get this type of 3D clone.

Techniques using deformable models can reconstruct 3D structured clones and eliminate noise depth data provided by RGB-D sensors. M.Zollhöfer et al [21] present an algorithm for 3D clones from high resolution RGB and depth data obtained with a Kinect camera. First, they smooth depth frame (Gaussian filter), detect feature points from the corresponding RGB image and segment the face using 3D depth. Then, they fit a Morphable 3D Face Model on the frame depth obtained by minimizing an energy term. Eventually, they project an RGB image on the 3D face reconstructed to obtain a 3D clone with a texture. C. Cao et al [2] create a database of 3D faces of 150 individuals. For each person, they capture with the RGB-D Kinect camera data from 15 different expressions included the neutral face. They then use Kinect fusion [17] to reconstruct the 3D face of each person. For each of these 3D faces, they detect 74 feature points using an Active Shape Models (ASM) [23] on the corresponding RGB images. Some points are manually adjusted for greater accuracy. Then, they fit a Morphable Face Model [10] on the 3D faces using an energy term. The model is deformed to adapt as effectively as possible 3D faces while matching characteristic points. Finally, they get a structured 3D mesh of each expression for each person. Note that their method is not fully automatic. M.Zollhöfer et al [14] presents an iterative method to clone the 3D face of a person. First, they detect the pose of the head and use a method similar to Kinect fusion to merge the different frames of depth and texture. Then, they detect the characteristic points of the face and they fit a statistical 3D face model [10] to reconstruct the shape and texture of the face. To do this, they optimize an energy term that finds the shape, albedo and illumination. They iterate these 4 steps for each new depth frame.

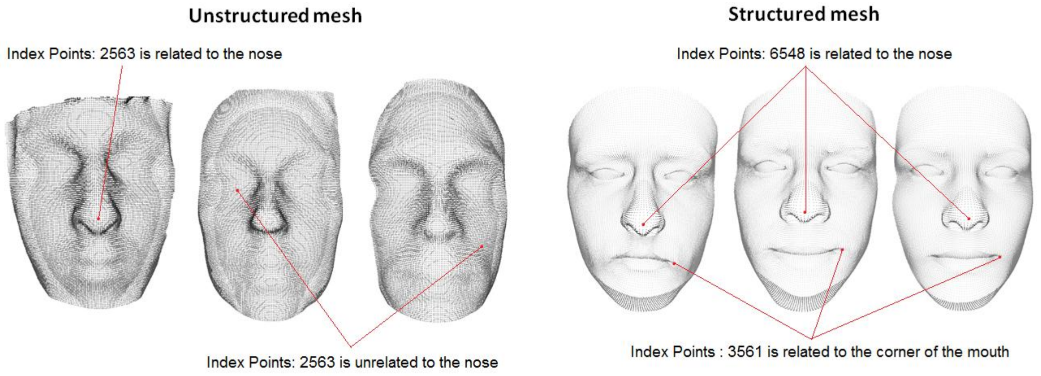
Unstructured mesh
Index Points: 2563 is related to the nose
Index Points: 2563 is unrelated to the nose
Structured mesh
Index Points: 6548 is related to the nose
Index Points : 3561 is related to the corner of the mouth

Figure 3: Definition of a structured mesh

The resulting 3D structured mesh obtained from these methods can be used in various applications such as a pre-process for gaze detection. Indeed, K.A.Funes Mora and J.Odobez [7] use 3D clones to estimate the pose of the head and the direction of the gaze of a person. To obtain a 3D clone, they fit a Morphable Face Model on data captured with a Kinect camera. Their method requires manual placement of feature points. They then use an algorithm based on Iterative Closest Point algorithm to detect the laying of the head. All the methods previously described above uses a Morphable 3D Face Model. They can provide high resolution structured clones. These techniques depend heavily on the quality of the model's face. Indeed, the specificities of the individuals can only be found if they belong to the database. It is therefore essential to use a database composed of diversified faces. The use of feature points can improve the fitting. But the methods are sensitive to the precision of the detection of these points. That is why most of these methods need manual adjustment of the points.

Our method of detection and fusion patches belongs to the category of techniques that use a Morphable Face Model but is fully automatic. Our technique is less dependent on the quality of the model used. Indeed, the goal of our algorithm is to improve the results of cloning using deformable models used in this type of methods.

3 Patches Detection and Fusion method

The different parts of our method are described in Figure 2. For each person, we capture various color and depth data of different views of the face. These data are noisy and in low resolution. We obtain a 3D point cloud (real coordinates: X, Y, Z and color: R, G, B) giving information about the subject's face. Our algorithm consists of two main sections: the reconstruction of the 3D shape (section 3.1) and the reconstruction of the 3D texture (section 3.2). For the reconstruction of the 3D shape, we use a Morphable Face Model. Compared to conventional methods, our process is reversed: we first perform a fitting with a Model on different depth frames Dp (p = 1 to n) to increase the resolution and remove noise and then we perform a fusion of the structured obtained meshes Mp (p = 1 to n) (Figure 1). We obtain a structured clone without texture Mc. In section 3.2, we describe the steps to rebuild the texture of the clone Tc. We explain how we map and merge the different texture images Ip (p = 1 to n).

3.1 Mesh Reconstruction

The first section of our process is the reconstruction of the 3D shape of the face. It is composed of two sub-sections: the fitting (3.1.1) and detecting and merging patches (3.1.2). The fitting and the patches detection are performed on each of the frames. Then we merge the obtained patches.

3.1.1 Fitting with a Morphable Face Model

The fitting is applied to each depth frame Dp (p = 1 to n). It is composed of a pretreatment and the iteration of two main stages. First, we perform preprocessing by filtering each depth frame Dp to remove part of the noise. Then, at each iteration, we align the depth frame Dp with the Morphable Face Model mesh S(α) (rigid transformation), and finally we deform the mesh S(α) so that it takes the shape of the depth frame Dp (non-rigid transformation). Each step is described below.

Bilateral Filter: Each depth frame Dp is smoothed with a bilateral filter before being treated [6]. This is a non-linear filter which has the advantage of preserving the edges and remove noise.

Rigid alignment: At each iteration of the fitting, we first need to align each depth frame Dp with the Morphable Face Model mesh S(α). The vector α is the parameter vector of the Morphable Face Model. We use the well-known and often used, iterative algorithm Iterative Closest Point [1], which aligns two 3D point clouds (rigid transformation). It consists of two stages: at each iteration, we match the points of $\mathrm{S}(\alpha_\mathrm{p})$ with the points of the depth frame Dp and then we estimate the rotation R and translation T matrix. Minimization of the error metric Eicp (equation 1) is used to estimate these two matrices.

$$\mathrm{E_{icp}(R,T) = arg\,min_E}\left\|\mathrm{S}(\alpha_\mathrm{p}) - (\mathrm{R} * \mathrm{D_p} - \mathrm{T})\right\|^2 \qquad \text{with p = 1 to n} \tag{1}$$

The error Eicp is based on the Euclidean distance between pairs of points in 3D point clouds that we want to align. Finally, these transformations R and t are applied to the frame Dp at the beginning of the next iteration. We iterate these two steps until the error Eicp reaches a minimum threshold or until the maximum number of iterations is reached. There are many variants of the ICP algorithm. S. Rusinkiewicz et al [22] compared the convergence characteristics of several ICP variants. For example, they used different distances (color, Euclidean ...) to match the points of two clouds to be aligned. In our method, we use the point to plan ICP of Y.Chen et al [3]. It is slower than the point to point but provides a better alignment of the two 3D point clouds. In the ICP algorithm, it is important to reject a maximum incorrect pairs of points. Therefore, we use a distance criterion to determine if a match is correct or not. We reject 50 percent of pairs of points.

Non-rigid transformation: After making the rigid transformation, we distort the average mesh $\bar{S}$ of the Morphable Face Model to fit to the depth frame Dp. In our process, we use Basel Face Model [10] to perform the non-rigid transformation. To compute this parametric model, they have made a principal component analysis (PCA) on 200 3D faces.

$$\mathrm{S}(\alpha) = \bar{\mathrm{S}} + \mathrm{U} * \mathrm{diag}(\sigma) * \alpha \tag{2}$$

In this equation, U is the orthonormal basis of the principal components of the PCA and σ the standard deviation of the components. The modification of the vector α provides the ability to distort the average face $\bar{S}$ to create a new 3D face. We compute a distance error Efit between the points of depth frame Dp and the mesh $S(\alpha_p)$ and we are looking for the α_p that minimizes this error:

$$E_{fit} = \mathrm{arg}\, min_E \left\| W_p .* \left(S(\alpha_p) - D_p\right)\right\|^2 \qquad \text{with p = 1 to n} \tag{3}$$

First, we match the mesh points $S(\alpha_p)$ and the depth frame Dp using the Euclidean distance between their points. It is important to eliminate incorrect pairs of points. Therefore, we use two criteria to reject incorrect matches based on the distance and the direction of normal vectors of the points. If the distance between the two points is greater than a preselected threshold and the angle of their normal vectors are not substantially identical, then we eliminate the pair of points. We calculate the error Efit using the weighted Euclidean distance Wp between pairs of selected points. Indeed, pairs of points with a short distance are the most important ones. That's why we use Wp weight inversely proportional to the distance between the matched points. Finally, we seek to change the coefficients vector α to find the minimum error Efit using a least squares optimization. At each iteration, the error Efit is recalculated. Figure 4 shows the evolution of the error on each vertex between the depth frame Dp and the average mesh $S(\alpha_p)$ at several iterations. Each depth frame contains various information about the face. A depth frame front view does not have any information on the profiles and on the sides of the nose. That is why in the figure 4 the error is greater in some parts of face (in red).

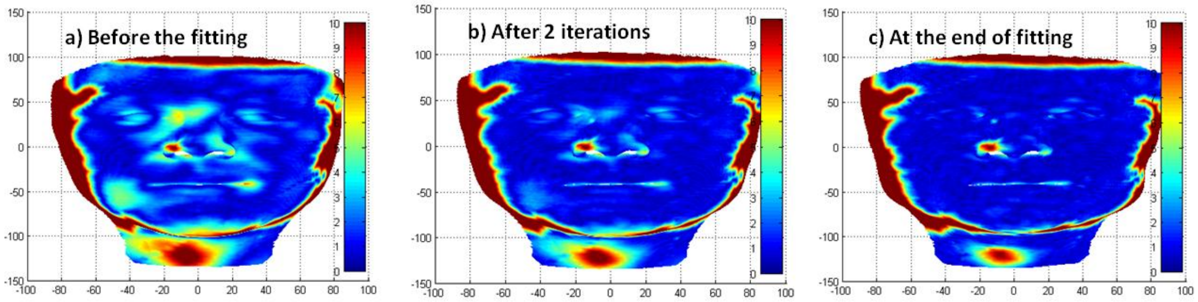

Figure 4: Evolution of error during non-rigid transformation

3.1.2 Patches Detection and Fusion on the shape

After making the fitting on each of the depth frames Dp, we obtain several structured meshes Mp. The aim of this second part is to detect the locations of the structured meshes Mp that are adequate: these places are called "patches". Then we merge these patches to create a mesh Mc.

Patches Detection: A structured mesh Mp which was created from a depth frame in right profile does not recover information of the left profile of the person as shown in Figure 5. For this reason, we want to keep only the parts of meshes that are adequate and accurate. For example, for a depth frame in right profile we want to keep the mesh patch that matches the right profile. We call "patch" all isolated points of each mesh we want to keep. Camera RGB-D captures more precisely the zones where the optical axis is perpendicular to the surface object. Therefore, we use a double condition. For a point to be preserved it must have a normal vector parallel to the optical axis of the camera and the distance between the mesh Mp and the depth frame Dp have to be smaller than a threshold. The value of this threshold is used to modify the precision of patches. Thus we get a patch for each mesh Mp. In Figure 5, we see two examples of patch detection on the 3D shape. We note that for a depth frame front view, we do not obtain all the information of the nose. Indeed, the error is large on the sides of the nose. For a frame in left profile, the error is very large to the right side of the face but little on the left side of the nose. The two depth frames of the figure 5 give different information.

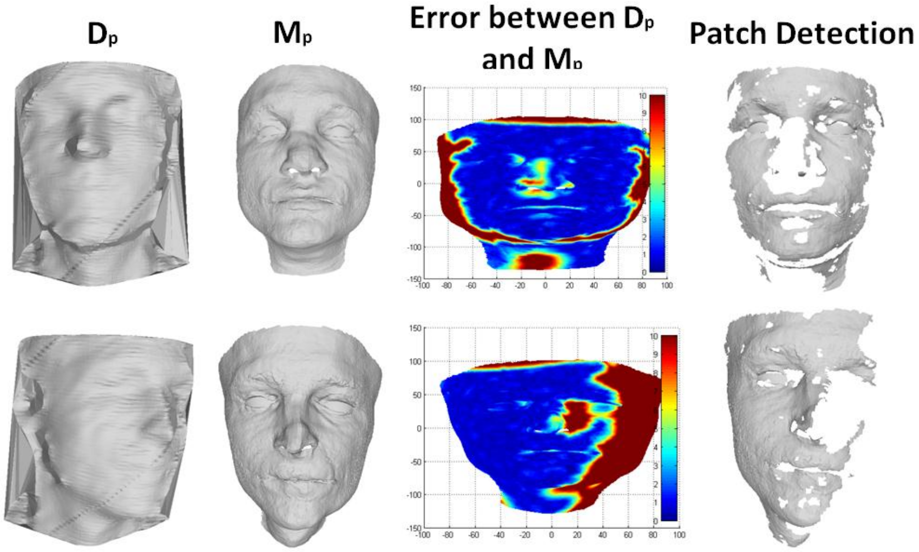

Figure 5: Example of patches detection on the shape.

Patches fusion: We want to merge the different patches we have detected to generate a complete 3D clone Mc (see Figure 2). All meshes Mp are structured. Therefore, we know the exact position of each patch on the face (eyes, forehead ...). For each point of the clone, there may be several overlapping patches (As the forehead of the face in the figure 5). That is why we make a fusion of points of these patches. We tested four types of fusion: the average, median, weighted average and robust average. For the average, we perform the average of overlapping points. For the median, we keep the midpoint. When there are not enough points (less than 3), the use of the median is not relevant. This is why we use the average value in this case. For the weighted average, the weights are the distances calculated in the step of fitting (section 3.1.1). For the fourth type of merger (robust average), we do not take into account the outliers in the calculation of the average. We eliminate the points that are away from the median value with a threshold (2mm). Figure 6 shows the result obtained with several depth frames Dp.

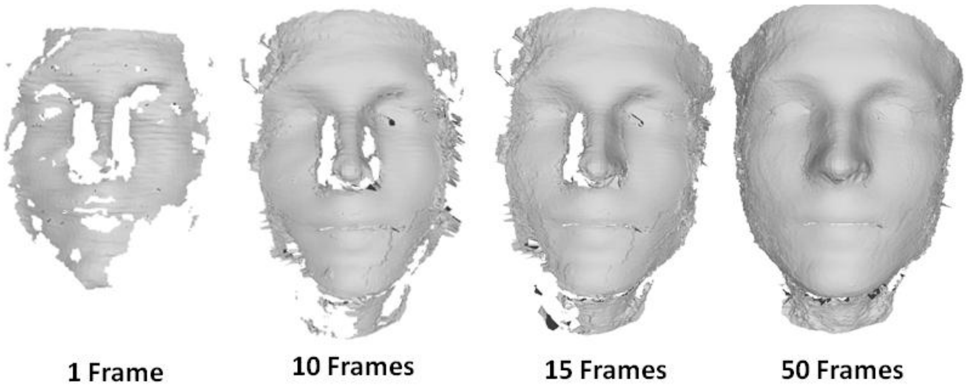

Figure 6: Patches Fusion on 3D shape (weighted average fusion)

At the end of this first step, we reconstructed the 3D shape of the face. We obtained a MC clone which will be used in the following step in order to reconstruct the texture (Figure 2).

3.2 Texture reconstruction

The second part of our method is the reconstruction of the texture. We use the same process as for the sub-section 3.1.2. This step is composed of two sub-sections: the texture mapping (3.2.1) and the patches detection and fusion on the texture (3.2.2).

3.2.1 Texture mapping

In this sub-section, we map the texture images Ip (p = 1 to n) on the structured Mc clone (Figure 2). We want several clones Tp (p = 1 to n) with **n** different textures Ip. Figure 7 shows an example of mapping for two frames, a frame profile view and frame front view. First, we align with the Mc clone each depth frame Dp using the ICP algorithm (described in paragraph 3.1.1). Camera RGB-D provides the mapping between the texture Ip and depth Dp. Then we map the textures Ip on the Mc clone using just this correspondence: for each vertex of the clone Mc, we map the texture corresponding to the closest point of the depth frame Dp. So we have several clones Tp with different textures.

3.2.3 Texture Patches Detection and Fusion

This sub-section consists of two stages: patches detection and patches fusion on the texture. We use the same procedure as for the 3D shape (see paragraph 3.1.2).

Patches Detection: We want to detect on each clone Tp, the texture patch that are adequate. Indeed, a Tp clone that was created from a texture image profile Ip does not recover the texture information on the left profile of the person as shown in Figure 7. To find out which texture points are relevant, we use the 3D shape of clones Tp and depth frames Dp. As in Section 3.1.2, we use two conditions: error distance between the clone Tp and the depth frame Dp and direction of normal vectors of their points. In Figure 7, we see two examples of patch detection on the texture. We note that for a left profile image Ip, we are not recovering the texture of the right profile. We observe that the shape error between the frame Dp and the clones Tp is relevant.

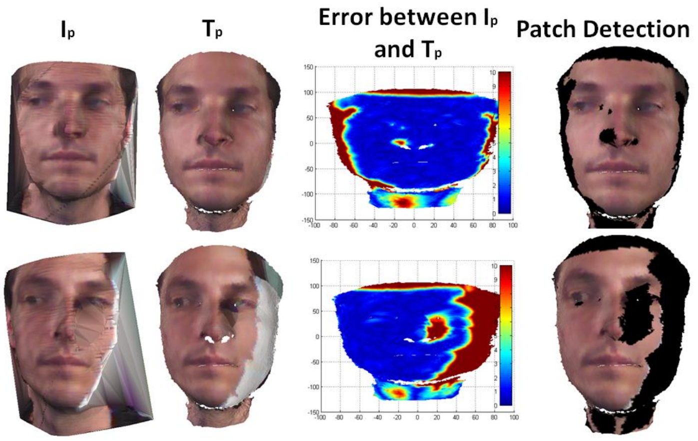

Figure 7: Example of patches detection on the texture.

Patches fusion: In this sub-step, we merge the patches detected in the previous step. We use the same method as in Section 3.1.2 to merge texture patches (RGB color of each point of the patches). We always use structured meshes Tp, which allows us to make the point to point fusion. We merge these patches for a complete facial texture. We also compare four types of fusion: average, median, weighted average and robust average. Figure 8 shows that the melting texture patches have been obtained with several images Ip using the median fusion.

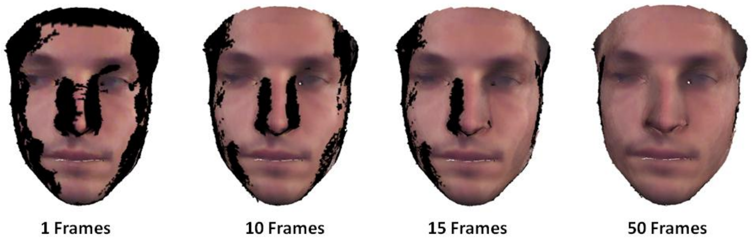

Figure 8: Patch Fusion on texture (median fusion)

Finally, our method allows reconstructing the 3D shape and texture of the face. We get a TC clone (Figure 2).

4 Experimental results

In this section, we first present the tools we have used (4.1 and 4.2) as well as our acquisition protocol (Section 4.3), then our results. We tested our method using Basel Face Model [10] and a Kinect camera. We compare our results qualitatively with other methods in the literature (section 4.5.1) and quantitatively using a ground truth (section 4.5.2).

4.1 Experimental protocol

We use a Kinect camera version 1 which is equipped with a color sensor and a depth sensor. It offers a resolution of 480 * 640 to 30 fps and it has a range at 0.5 meters. It does not work on reflective surfaces (the pupil) and in the presence of sunlight. For the fitting, we use the Basel Face Model (BFM) [10]. It was created from a training set of 200 scans of faces (100 women and 100 men). Each scan has a high resolution of 53,490 vertices (face and profile). The shape is statistically modeled by principal component analysis. Our acquisition protocol is simple and fast. Acquisitions are performed in a room with an ambient light. The subject performs a rotational movement of the head in front of the camera at 0.5 meter. He must do a neutral expression during the acquisition of data. For each person, the Kinect capture the texture Ip and the depth Dp. Our database of test consists of 6 subjects (Figure 13).

4.2 Comparisons of different fusions

For the reconstruction of the shape, we compared four methods of fusion: average, weighted mean, median and robust average. Figure 9 shows the clones obtained with the different methods for one subject. The method that uses the average does not eliminate outliers (artifacts) and the clone contains a lot of noise. The weighted average gives more importance to points of patches that are supposed to be correct, that are why it improves the results. With the median we get a better result. Indeed, one can see that there is less noise in areas without contour (cheek, forehead). However, we note that the eyebrows are more smoothed. We get the best results using the robust average. In fact, it eliminates a lot of noise (artifact) keeping details of faces: it removes the noise on the parts of the face without contour (cheek, forehead) while keeping the contour information (eyebrows ...). Indeed, it does not take into account certain items which are noisy patches in calculating the average.

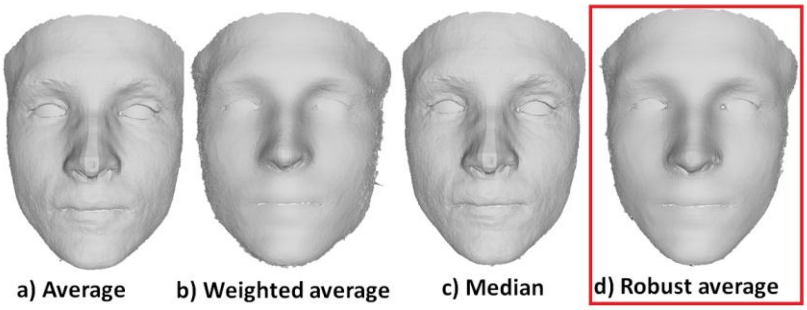

Figure 9: Comparison of different Patch Fusion on the texture

For the reconstruction of the texture, we also tested these four methods of fusion (Figure 10). We obtain a blur image of the texture when we use the average and the weighted average. The robust average slightly improves the results. Fusion with a median eliminates blur part of the texture and gives the best results. The low resolution of Kinect sensor (480 * 640) does not provide a high quality texture.

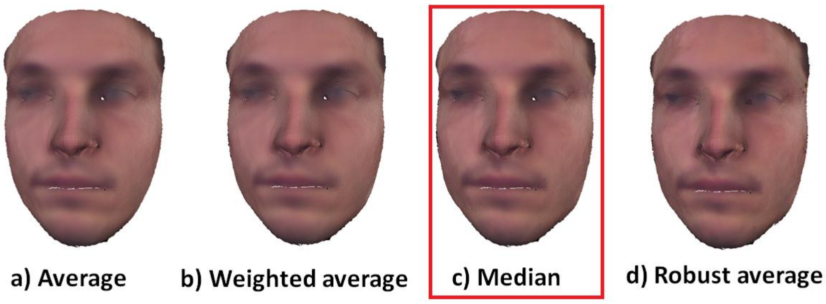

Figure 10: Comparison of several data fusion methods.

4.5 Results comparisons

We compared qualitatively (4.3.1) and quantitatively (4.3.2) our method with other methods in the literature. The qualitative comparison is used to compare the realism of the different results: we compared our results with the results of Kinect Fusion [17] and the FaceWarehouse process [2]. The FaceWarehouse process that we used consists of two steps: the merger of depth frames using Kinect Fusion and the fitting using the Basel Face Model (BFM) [10]. For fitting, we did not use the error term calculated with the feature points and the regularization term. Indeed, it is necessary that the detection of the feature points is very precise for the use of this error term is relevant. Methods for detecting the fully automatic feature points do not seem quite efficient (Active Shape Models (ASM) [23]…). Moreover, we do not use that regularization term because the algorithm converges without this term and the results obtained have fewer physical characteristics of the face. The quantitative comparison allows to know the accuracy of the results: we compared the results obtained with the FaceWarehouse process [2] and those obtained with our method with a ground truth.

4.3.1. Qualitative comparison

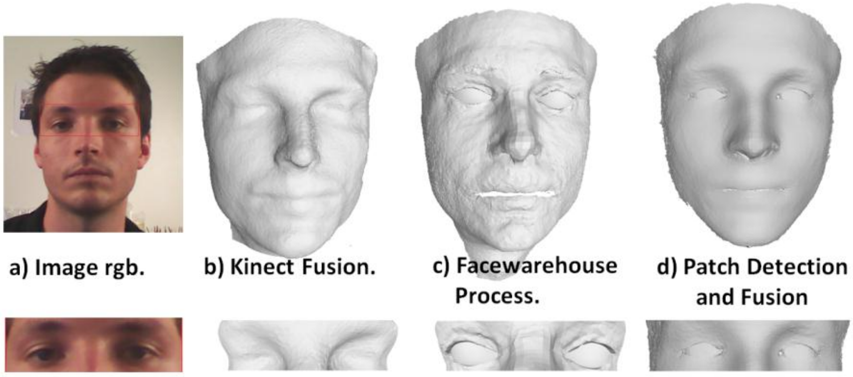

Figure 11: Qualitative comparison our results

Figure 11 shows the rendering of Kinect Fusion, FaceWarehouse process and our method on one subject. Kinect Fusion provides a clone with the specifics of the individual, but the facial features are not particularly strongly marked for instance at eye level. Ocular lobe does not appear on the 3D clone because the Kinect camera does not capture well the depth at eye level. Infrared rays are not efficient on the surfaces that reflect light (mirror, eyes ...). In addition, Kinect Fusion does not give a structured mesh. So it cannot be used directly in an application of gaze detection type [7] or facial animation for example. The FaceWarehouse process provides a structured clone where the facial features are pronounced. For example, the ocular lobe is not realistic. The clone obtained with the FaceWarehouse process has less of specifics of the individual. Indeed, Morphable Face Models are global models. In addition, they do not contain all possible forms and details of the subject's face and their learning databases are limited (200 faces for Basel Face Model). Our method is a compromise between the two previous methods. It provides that the facial features are well marked while also having more specifics of individual that a 3D clone created by FaceWarehouse process. For example, we can see that the eyes are smaller than obtained with the FaceWarehouse process. Hence, eyes of our 3D clone are more realistic.

4.3.2. Quantitative comparison

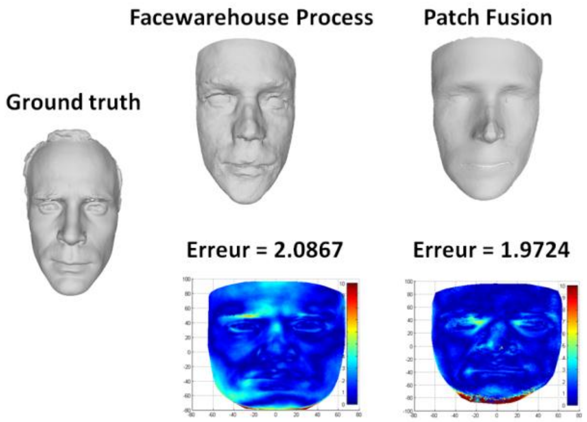

Figure 12: Quantitative comparison with ground truth

We also compared the results of our method and FaceWarehouse process with a ground truth. We do not make this comparison for our entire test database because we have the ground truth for only one subject. We did not calculate the error at eye level because the ground truth is not correct on this face area. First, we matched each point of the clone that we want to compare with the closest to the ground truth points. Then we calculated the overall error of distance between pairs of points (Figure 12). This figure shows the local error distance between each point of the two clones and the closest to the ground truth points. We can observe that the error is smaller with our method especially at the forehead and the chin. The error is larger at the level eye of 3D clone because the eyeball does not appear clearly on the ground truth. We observe that the overall error is smaller with our method (Error: 1.97) than with FaceWarehouse process (Error = 2.08). Figure 13 shows the results of our method on six subjects. It gives consistent results for different subjects.

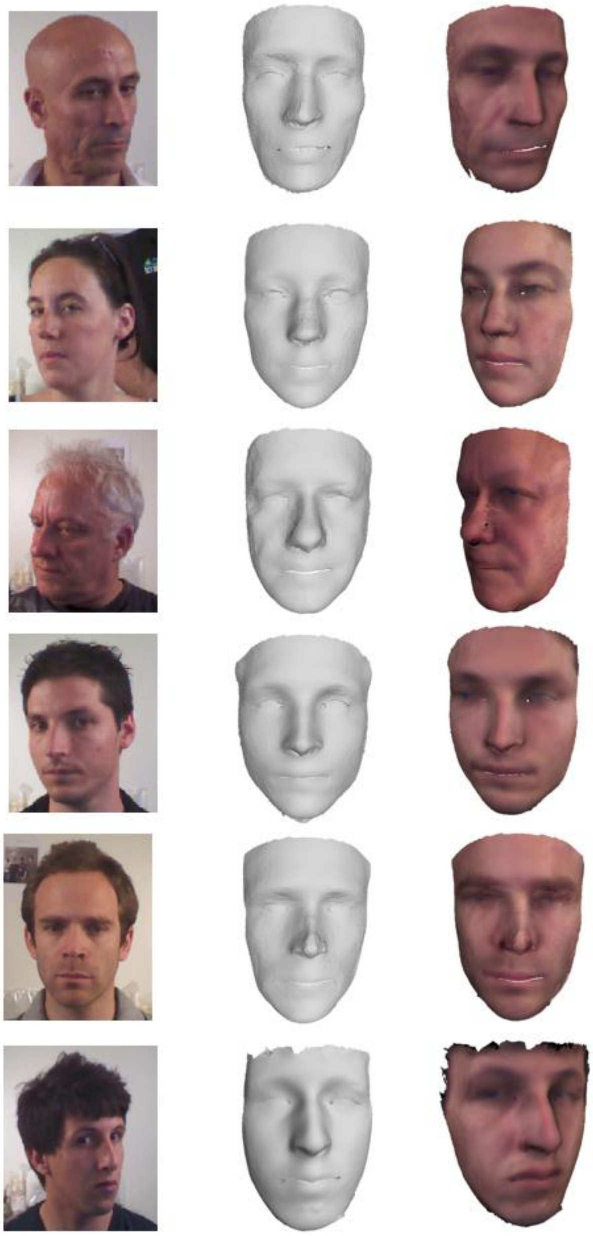

Figure 13: Results of our method on 6 people.

5 Conclusion

Face cloning is used in the field of video games and Human-Computer Interaction. Some applications require a system with low cost and easily accessible. Our method allows cloning faces with a low-cost sensor. We use a Morphable Face Model that allows obtaining structured 3D clones. The two contributions of our method are inversion process (fitting and fusion) and the use of shape and texture patches. We reverse the process to be less dependent on the alignment and assembly error. The use of patches makes it easier to find the specifics of an individual's face. We also observed that the reconstruction of the texture of the eyes is not correct. Therefore, we want to work in the future on the improvement of the quality of the texture. Using an interpolation could improve the results obtained with our method of texture mapping. In our future work, we also want to use the super resolution methods to increase the resolution of the data Kinect and improve the quality of our results. Finally, we wish to work on the cloning of facial expressions.

ACKNOWLEDGEMENTS

This research has been conducted with the support of Miles (FUI project) and REPLICA (ANR Techsan).

REFERENCES

[1]. Besl, P., McKay, N.D., 1992. *A method for registration of 3-d shapes*. Pattern Analysis and Machine Intelligence, IEEE Transactions on 14, 239–256.doi:10.1109/34.121791.

[2]. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K., 2014. *Facewarehouse: A 3d facial expression database for visual computing.* IEEE Transactions on Visualization and Computer Graphics 20, 413–425. URL: http://dx.doi.org/10.1109/TVCG.2013.249, doi:10.1109/TVCG.2013.249.

[3]. Chen, Y., Medioni, G., 1992. *Object modelling by registration of multiple range images.* Image Vision Comput. 10, 145–155.URL: http://dx.doi.org/10.1016/0262-8856(92)90066-C,doi:10.1016/0262-8856(92)90066-C.

[4]. Cui, Y., Schuon, S., Thrun, S., Stricker, D., Theobalt, C., 2013. *Algorithms for 3d shape scanning with a depth camera.* Pattern Analysis and Machine Intelligence, IEEE Transactions on 35, 1039–1050. doi:10.1109/TPAMI. 2012.190.

[5]. Debevec, P., Hawkins, T., Tchou, C., Duiker, H.P., Sarokin, W., Sagar, M., 2000. *Acquiring the reflectance field of a human face*, in: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques,ACM Press/Addison-Wesley Publishing Co., New York, NY, USA.pp. 145–156. URL: http://dx.doi.org/10.1145/344779.344855,doi:10.1145/344779.344855.

[6]. Fleishman, S., Drori, I., Cohen-Or, D., 2003. *Bilateral mesh denoising*. ACM Trans. Graph. 22, 950–953. URL: http://doi.acm.org/10.1145/882262.882368, doi:10.1145/882262.882368.

[7]. Funes Mora, K.A., Odobez, J., "Gaze estimation from multimodal Kinect data," Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on , vol., no., pp.25,30, 16-21 June 2012

[8]. Ralph Gross, Iain Matthews, and Simon Baker, "Generic vs. person specific active appearance models," Image and Vision Computing, Vol. 23, No. 11, November, 2005, pp. 1080-1093.

[9]. Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. 2008. Bosphorus Database for 3D Face Analysis. In *Biometrics and Identity Management*, Ben Schouten, Niels Christian Juul, Andrzej Drygajlo, and Massimo Tistarelli (Eds.). Lecture Notes In Computer Science, Vol. 5372. Springer-Verlag, Berlin, Heidelberg 47-56.

[10]. Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. 2009. A 3D Face Model for Pose and Illumination Invariant Face Recognition. In *Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance* (AVSS '09). IEEE Computer Society, Washington, DC, USA, 296-301.

[11]. Catherine Soladié, Nicolas Stoiber, and Renaud Séguier. 2013. Invariant representation of facial expressions for blended expression recognition on unknown subjects. *Comput. Vis. Image Underst.* 117, 11 (November 2013), 1598-1609.

[12]. Väljamäe, A., Larsson, P., Västfjäll, D. och Kleiner, M. (2004) Auditory Presence, Individualized Head-Related Transfer Functions, and Illusory Ego-Motion in Virtual Environments.

[13]. Jun Wang; Lijun Yin; Xiaozhou Wei; Yi Sun, "3D Facial Expression Recognition Based on Primitive Surface Feature Distribution," Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on , vol.2, no., pp.1399,1406, 2006 doi: 10.1109/CVPR.2006.

[14]. Zollhöfer, M., Thies, J., Colaianni, M., Stamminger, M., Greiner, G., 2014. Interactive model-based reconstruction of the human head using an rgb-d sensor. Journal of Visualization and Computer Animation 25, 213–222.

[15]. Sun, Q., Tang, Y., Hu, P., Peng, J., 2012. Kinect-based automatic 3d high resolution face modeling, in: Image Analysis and Signal Processing (IASP), 2012 Inter

[16]. Blanz, V.; Scherbaum, K.; Seidel, H.-P., "Fitting a Morphable Model to 3D Scans of Faces," Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on , vol., no., pp.1,8, 14-21 Oct. 2007

[17]. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S., Fitzgibbon, A., 2011. Kinectfusion: Real-time dense surface mapping and tracking, in: Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality, IEEE Computer Society,Washington, DC, USA. pp. 127–136.

[18]. B. Curless and M. Levoy. A volumetric method for building complex models from range images. ACM Trans. Graph., 1996

[19]. S. Schuon, C. Theobalt, J. Davis, and S. Thrun. “Lidarboost: Depth super-resolution for toF 3D shape scanning,” In CVPR, pp. 343-350, 2009.

[20]. Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt, “3D Shape Scanning with a Time-of-Flight Camera,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1173-1180, 2010.

[21]. Michael Zollhöfer, Michael Martinek, Günther Greiner, Marc Stamminger, and Jochen Süßmuth. 2011. Automatic reconstruction of personalized avatars from 3D face scans. Comput. Animat. Virtual Worlds 22, 2-3 (April 2011), 195-202.

[22]. Rusinkiewicz, S., Levoy, M., 2001. Efficient variants of the ICP algorithm, in: Third International Conference on 3D Digital Imaging and Modeling (3DIM).

[23]. Cootes, T. F. and Taylor, C. J. and Cooper, D. H. and Graham, J., 1995. Active Shape Models-Theirs Training and Application. Comput. Vis.Image Underst.