An Adaptive Approach to Real-Time 3D Non-Rigid Registration

Antonio Souza, Antonio Apolinário
Department of Computer Science
Federal University of Bahia
Salvador, BA, Brazil
Email: acsantossouza@gmail.com, apolinario@dcc.ufba.br

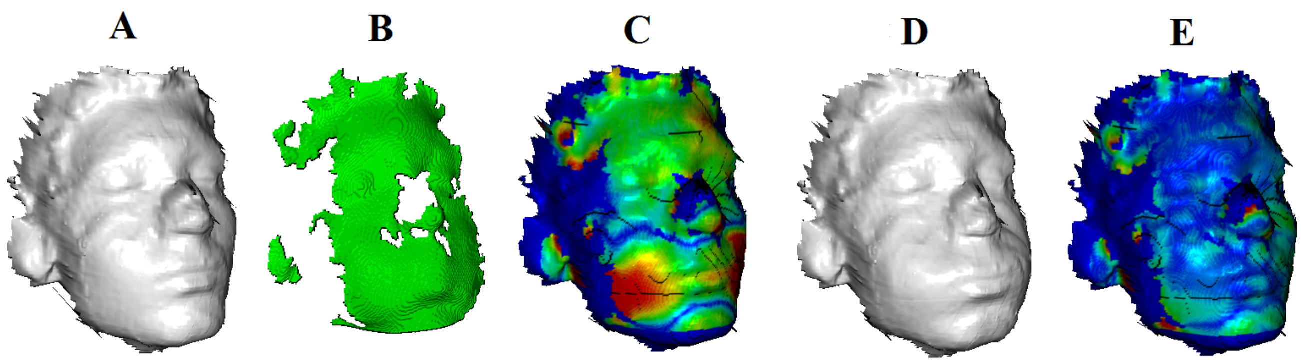

Fig. 1. Given reference (A) and deformed (B) surfaces related by a deformation error (C), our GPU-based adaptive non-rigid registration algorithm successfully captures the main deformation present on the deformed surface (D) by minimizing the initial error estimated (E) also running at interactive frame rates.

***Abstract*—3D non-rigid registration is fundamental for tracking or reconstruction of 3D deformable shapes. Fast methods for 3D non-rigid registration are particularly interesting for markerless augmented reality applications, in which the object being used as a natural marker may undergo a non-rigid user interaction. Here, we present a multi-frame adaptive algorithm for 3D non-rigid registration implemented on GPU where the 3D data is captured from an RGB-D camera. In this work[1], adaptivity is used in three stages of the process. First, to guide the distribution of regions of influence based on the deformation intensity on some region of the shape. Second, during the selection of constraints, where the sampling done over the object for the optimization is based on the current deformation. Third, to apply the algorithm in a multi-frame manner only when the rigid tracking error is above a pre-defined threshold. The results obtained show that the proposed algorithm is capable to achieve real-time performance with an approach as accurate as the ones proposed in the literature.**

***Keywords*-Adaptive Algorithms; Augmented Reality; Non-Rigid Registration;**

I. Introduction

Augmented Reality (AR) is a technology in which the view of a real scene is augmented with additional virtual information [1]. Tracking is one of the main problems which limits the development of a successful AR application because virtual and real worlds must be properly aligned so that they seem to coexist at the same location for the user.

In scenarios such as on-patient craniofacial medical data visualization [2], [3], it is specially important for a markerless AR (MAR) environment to provide support for non-rigid tracking, which adds one level of interactivity for the user and improves the robustness of the tracking algorithm for rigid and non-rigid patient interactions. The main issue related to this support is that AR requires real-time interactivity and most of the current state-of-the-art work in the field of 3D non-rigid registration does not provide such performance.

As will be shown in the next section, several approaches exist for accurate 3D non-rigid registration, however a few of them allow interactive registration. Despite the real-time techniques which rely on strong priors about a specific scenario, a few methods have been proposed for fast general-purpose non-rigid registration.

In this paper, we describe how to address the problem of fast 3D non-rigid registration by applying adaptive techniques to reduce the computational cost of the registration while keeping it accurate.

Contributions: The main contributions of our approach are the proposition of:

- A markerless augmented reality environment based on a low-cost RGB-D sensor;
- An adaptive algorithm for non-rigid registration;

[1]Ph.D. Thesis

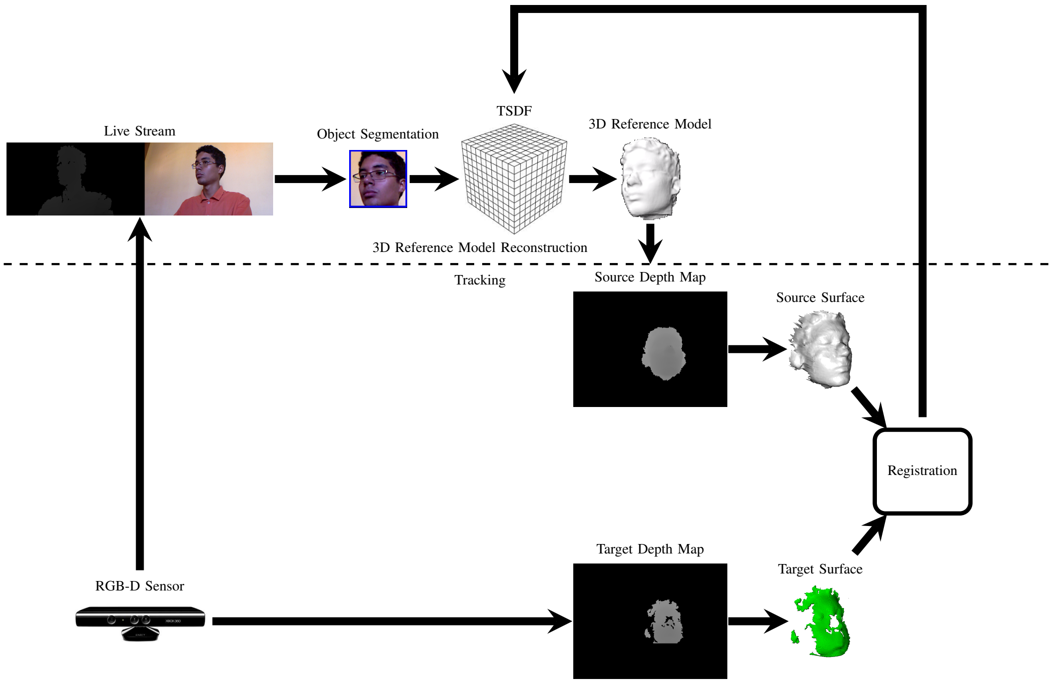

Fig. 2. Overview of the proposed approach from 3D reference model reconstruction to tracking solution.

- A full framework for non-rigid registration implemented entirely on the Graphics Processing Unit (GPU);

The rest of this work is organized as follows: Section II briefly describes the main methods proposed for interactive 3D non-rigid registration. Section III presents our first contribution, the MAR environment, in which the non-rigid registration is evaluated. Section IV describes our main contribution, the adaptive non-rigid registration, which uses GPU processing to achieve interactive frame rate. Section V discusses the results achieved with the proposed approach. Finally, Section VI concludes the work discussing possibilities for future research.

II. Related work

Several approaches have been proposed in the literature to register two deformable objects accurately. However, their major drawback relies on performance, which does not allow their use for interactive applications.

One of the first works for fast non-rigid registration applied to computer graphics is the Embedded Deformation (ED), a real-time deformation algorithm for object manipulation and creation of 3D animation [4]. The goal of this technique is to allow a user intuitive surface editing while preserving surface's features. Deformation is represented by a graph. Each node of this graph is associated with an affine transformation that influences the deformation to the nearby space. The great advantage of this approach is that it can be applied to a wide range of objects, articulated or not.

The ED algorithm enables user object manipulation performing as a non-rigid registration algorithm in which source and target surfaces are the objects before and after user manipulation. In this sense, many other works ([5], [6], [7], [8]) have used or improved this approach to the specific problem of non-rigid surface registration, focusing on improving accuracy while decreasing application's performance.

Few methods were capable to achieve real-time performance in 3D non-rigid registration. Chen et al. proposed a method for non-rigid registration of skeletons captured from the user's body [9]. Nutti et al. proposed a method to track tumors based on the patient's body position that presumes a prior knowledge about the scenario [10]. Zollhöfer et al. proposed a method for real-time non-rigid registration of arbitrary meshes captured from the real scene [11]. Based on a hardware specialized for high-quality surface acquisition, their approach generates a 3D template model of the object of interest and uses a hierarchical non-rigid registration algorithm fully implemented on the GPU. The implementation runs at 30 frames per second (FPS) with high accuracy.

In this work, we present an approach also based on the ED algorithm which shares some characteristics of [11], such as no special configuration or prior knowledge of the object and GPU parallelism to achieve real-time performance. However, no special hardware is supposed to be used, on the contrary, our approach is based on a simple off-the-shelf RGB-D sensor, with noise and low accuracy. As proposed by Li et al. [6], we use an adaptive graph refinement to improve non-rigid registration accuracy. Differently from other approaches, the algorithm proposed here runs entirely on the GPU and is based on a quadtree which operates over the 2D projection of the object to be registered. Also, the main goal of our algorithm is to be incorporated in a MAR environment, as a tool to improve tracking of the deformable object.

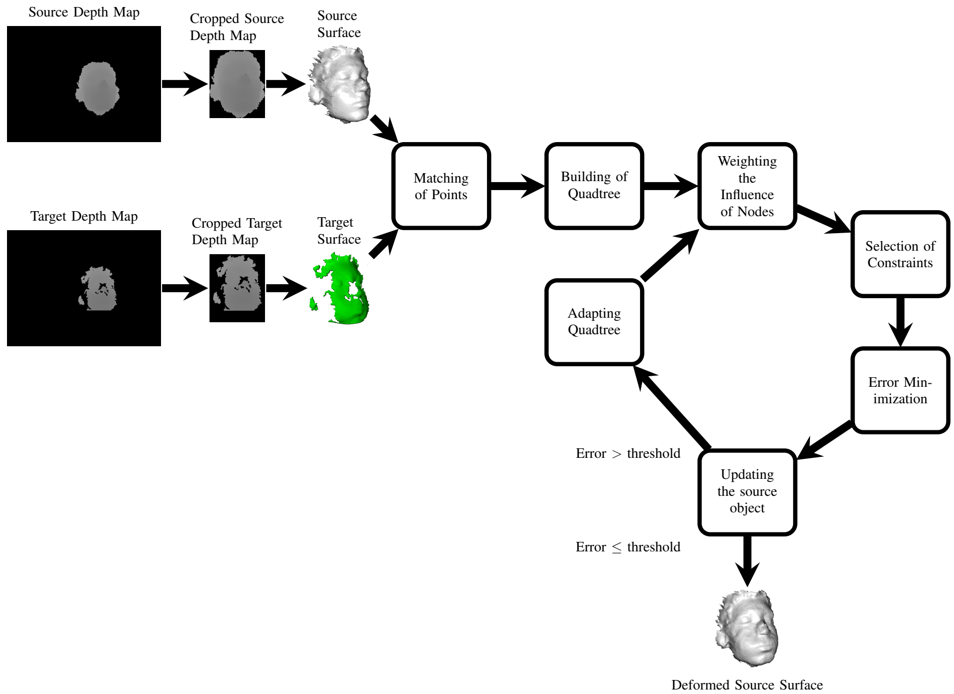

Fig. 3. Overview of the proposed approach from the depth map acquisition to the final non-rigid aligned surface.

III. Markerless Augmented Reality Environment

To validate the proposed real-time non-rigid registration, we have designed a MAR environment based on an RGB-D sensor (Fig. 2). Basically, a Microsoft Kinect sensor is used to capture color and depth information from the scene. Bilateral filter is employed to minimize the presence of noise preserving discontinuities of the data [12]. Moreover, unwanted points localized on the background scene are removed from the filtered depth map by using a depth threshold.

Then, the object of interest is localized and segmented from the scene by using a classifier or 2D bounding box which encloses the foreground object. After segmentation, the KinectFusion algorithm is used [13] to reconstruct a 3D reference model of the object of interest in real-time.

Real-time tracking is performed by using the 3D reference model previously reconstructed and the current 3D object captured by the sensor. The final registered 3D object is integrated into the 3D reference model to account for new viewpoints or changes in object's shape due to deformations. Rigid motion is estimated by a real-time variant of the ICP algorithm [14]. To find correspondences between the pairs of point clouds, it is used the projective data association algorithm, which matches the points that are located at the same 2D projection position (i.e. the same pixel). The ICP fails, or does not converge to a correct registration, when there is high pose variation between frames in sequence. To improve tracking robustness, a robust ICP based on real-time pose estimator is used to give a new initial guess to the tracking algorithm when it fails [15]. However, even using this algorithm, the tracking may fail if the user interacts non-rigidly with the application. Non-rigid tracking support can be added by applying a real-time non-rigid surface registration algorithm to align the 3D reference model and the current model captured.

To achieve real-time performance, all the steps of this MAR environment run on the GPU. All the algorithms were carefully designed and implemented in a parallel way to exploit the full parallelism provided by the hardware.

IV. GPU-Based Adaptive Non-Rigid Registration

The proposed non-rigid registration consists of several steps, illustrated in Fig. 3. Our deformation model is inspired in the algorithm proposed in [4]. However, we have added a three-level adaptive approach to improve accuracy and performance of the original solution. Moreover, we have implemented it on the GPU to boost performance.

The non-rigid registration algorithm builds a deformation graph (G) on the source point cloud (P_s) to allow its deformation to the target point cloud (P_t) iteratively. Each node $g \in G$ consists of a point $p_s \in P_s$ associated with a 3D affine rigid transformation (i.e. a 3D rotation matrix and a 3D translation vector) which influences the deformation to the nearby space. Our first adaptive solution is applied in the selection of nodes. Instead of using uniform sampling [4], we employ adaptive sampling in the GPU based on a quadtree. Basically, a quadtree is built on GPU based on the 2D projection of G. As the nodes of G are also points in P_s, we can convert them from world to image coordinates by using the same process used to project P_s in depth map. Based on the Euclidean distance measured between corresponding points of P_s and P_t, the nodes of G are refined or collapsed.

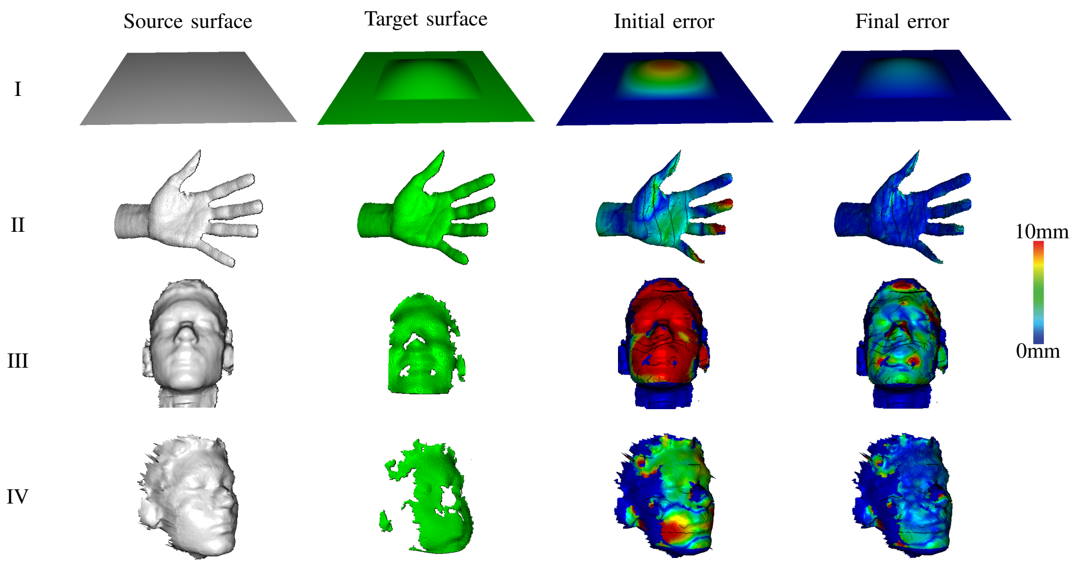

Fig. 4. The resulting color-coded error from the registration between source and target surfaces. In all situations the proposed algorithm obtained an averaged accuracy below $3mm$ and standard deviation below $3.5mm$.

Next, adaptivity is used to select the constraints, points from P_s which will be used for the optimization phase. This selection is based on the Euclidean distance measured between corresponding points. Given a region of P_s, the higher the error, the higher the number of points selected as constraints.

The last use of adaptivity is on the integration of the non-rigid registration algorithm into the MAR environment. Without adaptivity, one solution would be to apply the algorithm whenever the rigid tracking fails, enhancing the robustness of the MAR environment. However, to apply the non-rigid tracking for every frame has a computational cost which does not make it suitable for real-time applications. Therefore, if the rigid tracking keeps failing consecutively, the non-rigid tracking will be used more frequently, reducing user interactivity.

To solve this problem, we update the 3D reference model in real-time based on the current deformation measured. When the rigid tracking fails (i.e. the error measured is above a certain threshold), the non-rigid registration is applied and the 3D reference model is updated. As a consequence, by deforming the 3D reference model, the non-rigid tracking converges faster and with higher accuracy in the next iterations than the rigid-only solution (i.e. in which only rigid tracking is applied and the 3D reference model is not updated).

V. Results and Discussion

For all tests, we ran our algorithm on an Intel® Core™ i7-3770 CPU @3.50 GHZ, 8GB RAM memory, NVIDIA GeForce GTX 660. The 3D reference model was reconstructed with the KinectFusion using a grid with the volume size of $70cm \times 70cm \times 140cm$ and resolution of 512^3. Non-rigid registration takes $\approx$ 60ms per frame. Out of the MAR environment, we have tested the algorithm in four different datasets, as shown in Fig. 4. For the synthetic dataset I, the only deformation is the presence of a semi-sphere located on the center of the object. In this situation, our algorithm achieved high accuracy of $\approx 1mm$. For the dataset II, obtained from [16], a hand deforms by bending the fingers, where is the high error. The algorithm could reduce the average error below $2mm$. The datasets III and IV were captured by the Kinect sensor. The user was asked to smile and inflate his cheeks in front of the camera. In both cases, the algorithm had an accuracy of $2.6mm$.

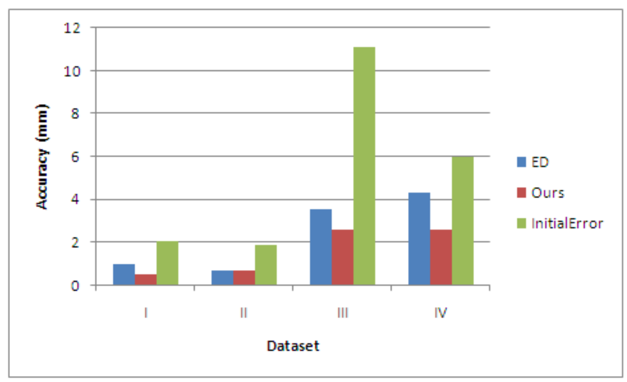

Fig. 5. Accuracy (in mm) obtained by our approach (Ours) in comparison with the Embedded Deformation (ED) algorithm and the initial error estimated for each one of the datasets used.

The improvement of accuracy by our approach with respect to related work is shown in Fig. 5. It occurs due to the adaptive selection of nodes, which redistributed the nodes in the deformation space, increasing them in the regions where the residual error is high and decreasing them otherwise. To improve even more the accuracy, one solution is to select more constraints to be used. Obviously, this decision will decrease the performance of the algorithm.

Related to performance, a comparison between algorithms can be seen in Fig. 6. As the graphic shows, our approach does not run in full real-time (i.e. 30 FPS), but it achieves interactive frame rates (i.e. above 15 FPS). The performance of each scenario varies mainly according to the size of the 3D models. Dataset I has $\approx$ 10k points, II has $\approx$ 80k points, III has $\approx$ 30k points and IV has $\approx$ 40k points. Nevertheless, the proposed non-rigid registration method is about three times faster than related work.

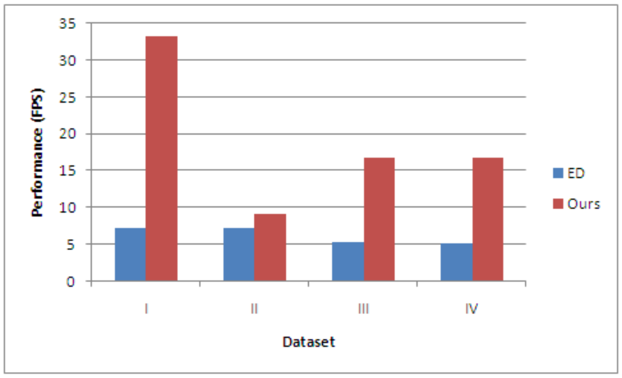

Fig. 6. Performance (in FPS) obtained by using our approach (Ours) in comparison with related work in the field of interactive non-rigid registration (ED - Embedded Deformation algorithm [4]) for each one of the datasets used.

The use of adaptivity for constraint selection greatly reduces the processing time originally demanded by the ED algorithm. Optimization is a common bottleneck in non-rigid registration algorithms [4], [6]. The number of constraints selected is directly related to the time required by the optimization phase. Therefore, by reducing adaptively the number of constraints used, we can achieve good performance even in the optimization phase. Moreover, as long as the error is minimized over the surface, the number of nodes is dynamically decreased from G. With less parameters to be estimated, the optimization algorithm converges faster.

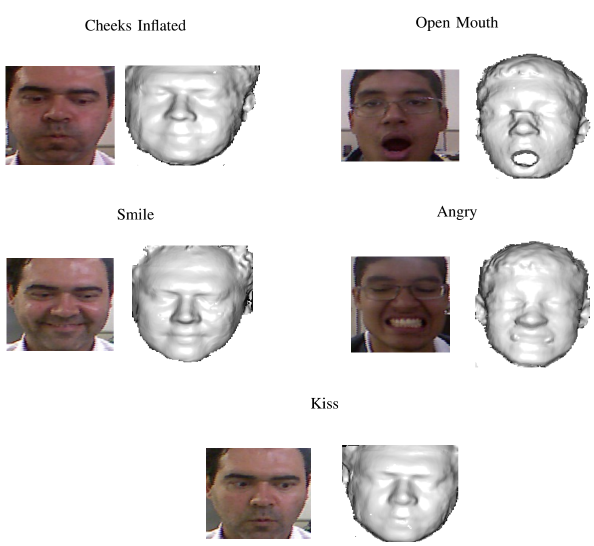

Fig. 7. Deformed reference models for different users and facial expressions.

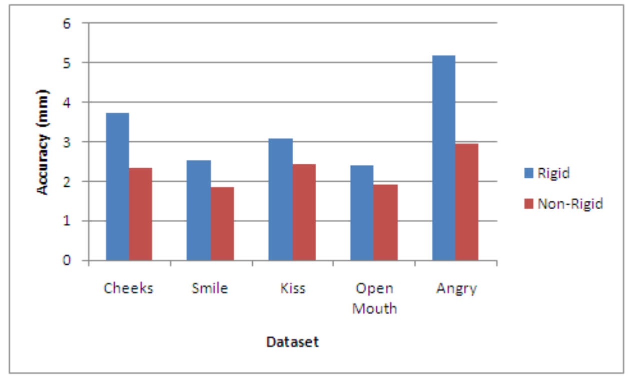

Fig. 8. Average accuracy (in mm) measured for each one of the facial expressions used to evaluate of our approach.

For the tests of our algorithm on the MAR environment, we have tested the approach in a scenario where the user's head is our main natural marker. Therefore, we asked two users to perform five different facial expressions: to inflate his cheeks, smile, kiss, open mouth and to be angry, as shown in Fig. 7. In these cases, we estimated an average accuracy of $1.5mm$ for rigid tracking during 3D rigid reference model reconstruction.

When non-rigid user interaction is present, the average accuracy decreases for rigid tracking. As can be seen in Fig. 8, to apply the non-rigid registration whenever the rigid tracking fails is a good idea in order to solve every deformation which occurs between frames.

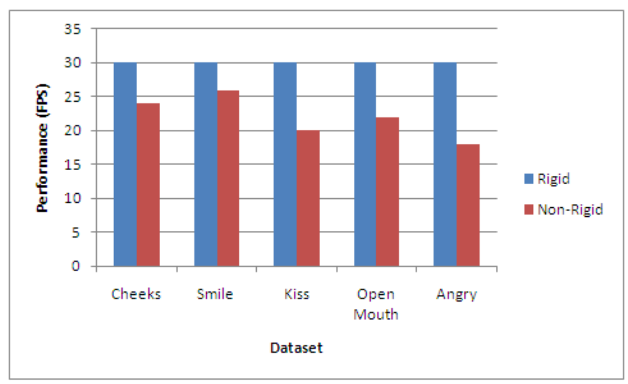

Fig. 9. Average accuracy (in FPS) measured for each one of the facial expressions used to evaluate of our approach.

From Fig. 9, it is visible that the proposed approach allows interactive performance. It is worthy to mention that, in this case, the algorithm is not applied almost for every frame as the 3D reference model is updated based on the current deformation, reducing the chances for rigid tracking failure in the next iterations.

VI. Conclusion and Future Work

In this work we have proposed a new adaptive non-rigid registration algorithm which tracks interactively and with high accuracy 3D deformable objects captured from the real scene in a MAR environment. As a result, we have designed a non-rigid registration technique which outperforms the performance of existing related work, while keeping moderate accuracy on the final solution. To achieve such conclusion, tests were conducted on four datasets which are characterized by their different levels of noise and precision. To validate the deformable registration in the context of the MAR environment, tracking robustness was evaluated for several challenging scenarios. The algorithm was applied in a multi-frame way, which allowed averaged accuracy between 2 and $3mm$, and real-time performance.

For future work, one can incorporate into our approach highly efficient optimization algorithms, such as [11]. As we could reduce the dimensionality of the optimization from the adaptive selection of nodes and constraints, it is expected for our solution to achieve even better performance than previous approaches even using low-cost depth sensors. On the current state, our approach does not support non-rigid registration between non-isometric structures (e.g. faces of different people, models which have grown). State-of-the-art solutions for accurate shape matching can solve the problem of correspondence between these structures, however, these methods are far from real-time performance.

VII. Related Publications

1) A. Souza, M. Macedo and A. Apolinario. "Multi-Frame Adaptive Non-Rigid Registration for Markerless Augmented Reality". in ACM VRCAI. **(Conference - Qualis: B2 - Oral Presentation)**, pp. 7-16, 2014.
2) A. Souza, M. Macedo and A. Apolinario. "A GPU-Based Adaptive Algorithm for Non-Rigid Surface Registration". in IEEE VR. **(Conference - Qualis: A2 - Poster Presentation)**, 2015.
3) M. Macedo, A. Apolinario, A. Souza. "A Robust Real-Time Face Tracking Using Head Pose Estimation for a Markerless AR System". in SVR. **(Conference - Qualis: B4 - Oral Presentation)** pp. 224-227, 2013.
4) M. Macedo, A. Apolinario, A. Souza. "KinectFusion for Faces: Real-Time 3D Face Tracking and Modeling Using a Kinect Camera for a Markerless AR System". in SBC Journal on 3D Interactive Systems. **(Journal - Impact Factor: 0.411)** vol. 4, pp. 2-7, 2013.
5) M. Macedo, A. Apolinario, A. Souza. "A Markerless Augmented Reality Approach Based on Real-Time 3D Reconstruction using Kinect". in SIBGRAPI. **(Conference - Qualis: B1 - Poster Presentation)**, 2013.
6) M. Macedo, A. Apolinario, A. Souza, G. Giraldi. "High-Quality On-Patient Medical Data Visualization in a Markerless Augmented Reality Environment". in SBC Journal on 3D Interactive Systems. **(Journal - Impact Factor: 0.411)** vol. 5, pp. 41-52, 2014.
7) M. Macedo, C. Almeida, A. Souza, J. Silva, A. Apolinario, G. Giraldi. "A Markerless Augmented Reality Environment for Medical Data Visualization". in WIM. **(Conference - Qualis: B4 - Oral Presentation)** pp. 1682-1691, 2014. **[Best Paper Award]**
8) M. Macedo, A. Apolinario, A. Souza, G. Giraldi. "A Semi-Automatic Markerless Augmented Reality Approach for On-Patient Volumetric Medical Data Visualization". in SVR. **(Conference - Qualis: B4 - Oral Presentation)** pp. 63-70, 2014.

Acknowledgment

This research is financially supported by FAPESB and CAPES.

References

[1] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, Aug. 1997.
[2] J.-D. Lee, C.-H. Huang, T.-C. Huang, H.-Y. Hsieh, and S.-T. Lee, "Medical augment reality using a markerless registration framework." *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5286–5294, 2012.
[3] M. Macedo, A. Lopes Apolinario Junior, A. Souza, and G. Giraldi, "A semi-automatic markerless augmented reality approach for on-patient volumetric medical data visualization," in *Virtual and Augmented Reality (SVR), 2014 XVI Symposium on*, May 2014, pp. 63–70.
[4] R. W. Sumner, J. Schmid, and M. Pauly, "Embedded deformation for shape manipulation," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007.
[5] H. Li, R. W. Sumner, and M. Pauly, "Global correspondence optimization for non-rigid registration of depth scans," *Computer Graphics Forum (Proc. SGP'08)*, vol. 27, no. 5, July 2008.
[6] H. Li, B. Adams, L. J. Guibas, and M. Pauly, "Robust single-view geometry and motion reconstruction," *ACM Transactions on Graphics (Proceedings SIGGRAPH Asia 2009)*, vol. 28, no. 5, December 2009.
[7] H. Li, L. Luo, D. Vlasic, P. Peers, J. Popović, M. Pauly, and S. Rusinkiewicz, "Temporally coherent completion of dynamic shapes," *ACM Transactions on Graphics*, vol. 31, no. 1, January 2012.
[8] M. Dou, H. Fuchs, and J.-M. Frahm, "Scanning and tracking dynamic objects with commodity depth cameras," ser. ISMAR '13. IEEE Computer Society, 2013.
[9] J. Chen, S. Izadi, and A. Fitzgibbon, "Kinetre: Animating the world with the human body," ser. UIST '12. New York, NY, USA: ACM, 2012, pp. 435–444.
[10] B. Nutti, A. Kronander, M. Nilsing, K. Maad, C. Svensson, and H. Li, "Depth sensor-based realtime tumor tracking for accurate radiation therapy," *Proc. of Eurographics 2014 Short Papers*, April 2014.
[11] M. Zollhöfer, M. Nießner, S. Izadi, C. Rehmann, C. Zach, M. Fisher, C. Wu, A. Fitzgibbon, C. Loop, C. Theobalt, and M. Stamminger, "Real-time non-rigid reconstruction using an rgb-d camera," *ACM Transactions on Graphics (TOG)*, 2014.
[12] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *ICCV*, jan 1998, pp. 839 –846.
[13] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera," ser. UIST '11. USA: ACM, 2011, pp. 559–568.
[14] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *3DIM*, Jun. 2001.
[15] M. Macedo, A. Apolinario, and A. Souza, "A robust real-time face tracking using head pose estimation for a markerless ar system," in *Virtual and Augmented Reality (SVR), 2013 XV Symposium on*, May 2013, pp. 224–227.
[16] T. Weise, B. Leibe, and L. Van Gool, "Fast 3d scanning with automatic motion compensation," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, June 2007, pp. 1–8.