Automatic Animation of High Resolution Images

Dmitry Batenkov, Gregory Dinkin, Yosef Yomdin
Department of Mathematics
The Weizmann Institute of Science
Rehovot 76100, Israel

Abstract—**In the report, we present a novel photo-animation method and an animation tool for transformation of a photo into an animated video clip. In opposite to the existing tools, our technology provides both automatic and intuitive interactive photo-animation of high-resolution images.**

Our method and technology comprise the following main components:

1. **Vectorization of the image, i.e. its representation by the sets of primitive structures (geometric models) in different scales.**
2. **3D-2D global scale kinematic models for various image objects, especially, for the patterns of human bodies. These models allow an adaptation to definite parameters of the patterns and provide a high visual quality of motions' reproduction, including relatively large 3D rotations.**
3. **Tools and methods for translation of true 3D motion (like various scenarios of human 3D motion appearing in the Carnegie Mellon University database) into 2.5D animation.**
4. **Tools and methods for real-time automatic fitting of the models to image objects. In the course of fitting, specific anatomic parameters of the pattern are detected and preserved. The suggested method starts with face detection and model fitting, and then extends up to a full body pose capturing and model fitting. Our approach strongly relies on information provided by a vectorized image data, like the geometry of the edges and ridges and their color profiles, and on the transformations of 3D and 2D models.**
5. **High level scenarios and their automatic adaptation to the actual position of the characters in the photo.**

The presented technology is illustrated by several examples. Links to more running examples and to our on-line Photo-animation tool are provided.

Index Terms—**Image processing, geometric modeling, photo-animation.**

I. INTRODUCTION

Photo-animation is a transformation of a static photo into a movie, and of the characters appearing in the photo into virtual actors. This process is intensively used on the "high-end" platforms for computer imaging and today requires a tedious work of professional animators.

There are several very difficult problems on the way to "easy photo-animation" intended to non-professional users and to the common low-end platforms: personal computers and smart-phones in their various versions.

The first problem is that in order to transform a photo-character into a virtual actor, it must be *automatically* accurately fitted with an appropriate parametric human 3D or 2D model. This problem has been addressed in many recent publications (see [3], [5], [10], [12], [13], [14] and references therein for a small sample of the literature). However, the existing solutions, even in the most studied case of the human face, do not provide a fitting accuracy required for the convincing animation. This problem becomes especially acute for characters in complicated poses, partially occluded (as it usually happens on album photos). So in order to make easy photo-animation feasible we need *automatic tools for a real time accurate fitting of human (and other) photo-characters with parametric models*.

Secondly, as for today and for a near future, the high quality *and especially, high resolution* 3D graphics cannot be used in real time applications on the low-end platforms. In order to preserve high resolution of the input photo we are forced to use a kind of "2.5D" animation format. There are several formats of such kind, including Adobe Flash [1] and AnimeStudio [2], but the resulting animations are usually cartoon-like, both in the motion and the texture quality. So *the problem is to develop a 2.5D animation format, which will provide high quality human motions and high-resolution texture and will be suitable for a playback on low-end platforms*.

The third problem is related to the fact that even if virtual actors, virtual environment, and easy control tools are available, animation remains to be highly nontrivial creative task. A non-professional user typically cannot produce anything beyond short isolated actor motions. Another, more technical problem is that the usual animation tools, like time-line, key frames, motion interpolation, require serious training and *understanding of the time structure of the animation*. None of these can be expected from an average Internet user. There are many publications treating this problem and, in particular, suggesting new animation interfaces and procedures (see, e.g., [6], [7], [11], [15]), and, certainly, this problem goes far beyond the task of photo-animation. However, at least two its small parts have to be addressed in order to make easy photo-animation feasible:

1. We need a *format and tools for "template animation scenarios", which can be automatically adapted to the actual number and position of human characters on the input photo*.
2. We need *interactive tools, which will give to the user certain control of a scenario*, at least in the limits preserving the basic time-motion structure of the animation. For example, the user would be able to add new elements compatible with

978-1-4673-4681-8/12/$31.00

the main animation, to control facial expressions, style and intensity of the motions, etc.

In this report we present our photo-animation method and an animation tool, stressing the solution of the first two problems, while our solution of the third problem (very partial at this moment) is shortly outlined.

Running examples can be found on the link http://photo-animator.weizmann.ac.il/info/geometricmodels.htm; on-line Photo-animation tool is http://photo-animator.weizmann.ac.il.

II. AUTOMATIC FITTING OF PARAMETRIC MODELS

We shall discuss fitting of *enface* human model shown in Figure 1. The specific feature of this model is that it is formally represented as a projection of standard 3D human model to the viewer's plane. Thus, 3D transformations of the 3D model's skeleton are analytically translated to the transformations of the skeleton and the layers of 2D model (see Section III).

The fitting algorithm has been implemented also for a *profile* human model, and for a detailed *human face* model.

The first stage of the algorithm consists of a rough identification of the photo-character and of a corresponding initialization of the model position. This can be achieved via algorithms described in [12] or similar ones. Figure 1 presents the input photo and the result of the initialization step.

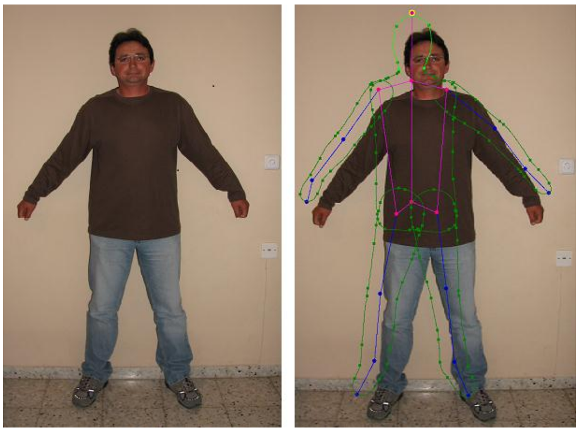


Fig.1. The input photo and the result of initialization step for model fitting.

At the second stage (of an accurate fitting) we conduct a multi-scale minimization of a distortion of the 2D model's contours from boundary edges of the photo-object.

Our approach relies on the information provided by a vectorized image data [8], [9], like geometry of the edges and ridges and their color profiles. This information is incorporated into a distortion function, which has to be minimized. Parameters of the 2D model include kinematic characteristics of both 2D and 3D skeletons. Accordingly, in addition to "direct" 2D parameters, 3D transformations, as they are reflected in the transformations of 2D models, are used as minimization parameters.

At this stage, serious difficulty has to be settled: the indicated minimization problem is highly non-linear, and the number of its free parameters is greater than two hundreds. Any attempt to solve such a problem in a straightforward way fails. Certainly, free parameters of the model can be separated into the kinematic parameters of 3D and 2D skeletons (in our case there are nearly 70 and 25 such parameters, respectively), and into control points of the contours (up to 200), which are responsible to the contours' fine fitting. So, minimization is naturally split into rough "skeleton minimization" and further fine "contours minimization".

However, skeleton minimization with 70 (or even 25) parameters, in any attempt of its straightforward solving, turns out to be unfeasible. To overcome this difficulty we suggest a procedure of "almost decoupling", which provides a robust real time solution. This procedure acts as follows.

Let S be the set of the skeleton parameters. It includes joints positions, and bones directions and lengths.

In "almost decoupling" we first define certain **groups** S_i of the active parameters, $i = 1, \dots, N$. Each group S_i may contain at most $3-4$ parameters, and to each such group we associate a certain part C_i of the model's contour (see Figure 2 and Figure 3). The groups S_i and contour paths C_i must be chosen in such a way that:

1. The problem of the distortion minimization on the part C_i of the model contour, with respect to the parameters S_i, has a unique and stable solution for each $i = 1, \dots, N$ ("rigidity").
2. The influence of the parameters S_i on the contours C_j for $i \neq j$ is relatively small ("almost decoupling").
3. Parameters' set S is covered (with possible overlaps) by the groups S_i for $i = 1, \dots, N$.

Next we conduct minimization for each of the groups S_i separately, one after another, from $i = 1$ to $i = N$, and repeat this minimization circle several times. It can be proven that (under a proper quantification of the requirements 1-3 above) the "almost decoupling" procedure converges to a true solution of the original minimization problem, while drastically reducing the computation load. We plan to present this mathematical result separately.

Practically, the groups S_i for our models have been found experimentally, in a rather long process of tuning by "trial and error". At present, we have around $N = 30$ groups, three of which are shown on Figure 2 and Figure 3. Figure 2 shows two of the groups S_5 and S_{11} for the skeleton fitting.

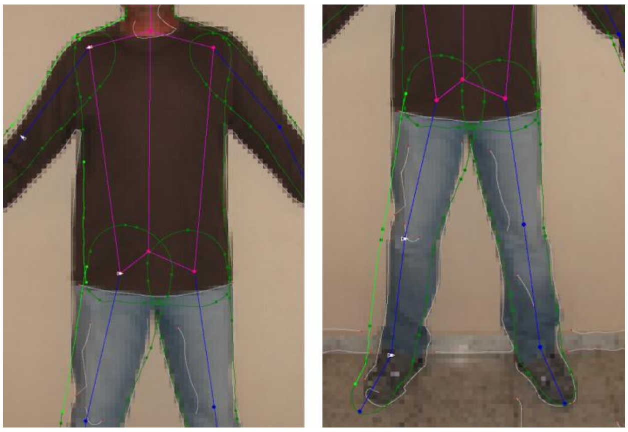


Fig.2. Two of the groups S_5 and S_{11} for skeleton fitting.

Figure 3 presents one of the “medium scale” groups S_{21} for the contours fitting. Following application of all the medium scale groups for the contours fitting, a fine scale fitting is applied, where the contours control points are just moved to the nearest point of the appropriate edge of the photo-object (as it was identified in the previous fitting stages).

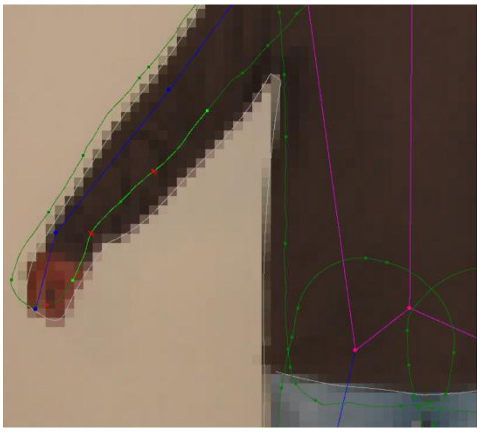


Fig.3. “Medium scale” group S_{21} for the contours fitting.

Assuming that the initialization error was of order of one fifth of the model size (relatively to the figure), our minimization procedure takes a fraction of a second and produces an accurate fitting, as shows Figure 4.

Fig.4. Result of the fitting procedure.

There is another difficult problem in model fitting which is settled with the “almost decoupling” procedure: it is easy to see that typically solution of skeleton minimization problem is not unique. Indeed, in many cases stretching some bones and shortening others beyond the anatomic proportions provides a better fitting of the model contours. However, such “non-anatomic” fitting will completely destroy the motions of the resulting virtual artist. The “almost decoupling” procedure prevents such effects, since fitting rigidity is required to hold inside each of the groups S_i. Such construction of the groups excludes “non-anatomical” model deformations form the fitting process.

III. 3D QUALITY FOR 2D MODELS ANIMATION

As it was explained above, in order to preserve high resolution of the input photo while playing animations on the low-end platforms, we are forced to develop a 2.5D animation format, which provides high quality of human motions and of high-resolution texture. To achieve this goal we combine three known graphics tools: full 3D human models, skeleton based 2D animation (as in AnimeStudio [2]), and a layer based 2.5D animation (as in Flash [1]).

Our full 3D human model may reproduce high quality 3D human motions in a standard way, as it is obtained using motion tracking or other types of 3D animation. There exist also rich data bases of human 3D motions, in particular, that of the Carnegie-Mellon University ([4]).

Currently, we use two main methods to transform a 3D motion of the 3D model into a high quality animation of the 2D model: skeleton 3D animation and layers-based 3D animation.

In the first method, the skeleton movements of the 2D model are computed as true projections to the “visual plane” of the 3D skeleton of the 3D model. Then, we apply one of the standard mathematical procedures of 2D skeleton animation, in particular, that of AnimeStudio [2]. For relatively small 3D motions, visual illusion of true 3D motion is obtained. Figure 5 illustrates this approach with an example of a human face 3D motion, which is obtained by skeleton animation method.

Fig.5. 3D-2D skeleton motion combined with skeleton 2D animation.

The second method is based on the 2.5D layers animation (as in Flash [1]). In this method, projective transformations of the layers, dynamic computation of their relative depth (and the corresponding z-buffering) are performed. Starting with a full 3D animation of the 3D model, we compute projective transformations of the layers of the 2D models, which approximate in the best way true 3D motions, as they are seen from a prescribed view direction. This is done in three steps:

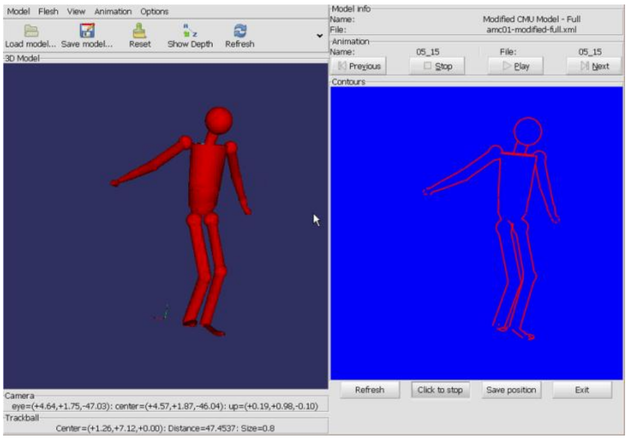
1. With the 3D skeleton of the 3D model we associate simplified analytic models of the corresponding limbs.

Typically, these are parabolic cylinders with elliptic sections, defined by the fourth degree equations.

2. For each position of the 3D skeleton we compute approximate analytic projections of the limbs to the viewer's plane.

3. Finally, we find projective transformations of the original layers, which best fit the limbs projections.

Figure 6 shows an example of 3D model and its analytic projection to the viewer's plane. Figure 7 presents the 2D model fitted to the photo character (on the left), and a frame of the resulting layer-based animation (on the right).



Name: Modified CMU Model - Full
File: amc01-modified-full.xml

Name: 05_15
File: 05_15

eye=(+4.64,+1.75,-47.03): center=(+4.57,+1.87,-46.04): up=(+0.19,+0.98,-0.10)

Center=(+1.26,+7.12,+0.00): Distance=47.4537: Size=0.8

Fig.6. 3D model and its analytic 2D projection to the viewer's plane.

Fig.7. Photo character and a frame of the resulting animation.

IV. Easy Animation Tools and Interfaces

As it was mentioned in the Introduction, to provide a photo-animation tool for non-professional users, we have to address at least two of several extremely difficult problems rising on the way to “easy animation”. The first is a problem of automatic adaptation of the template scenario to actual position of the actors on the input photo, and the second is providing the user with reasonable interactive controls of the animation. Our very partial solutions at this stage are the following:

1. The template animation scenario includes parameters of actual position and shape of virtual actors on the photo. The motions of the actors and objects are automatically recalculated according to these parameters.

2. We are developing interactive tools, which allow the user to control the animation without entering the “forbidden area” of time line and actors interactions. The obtained solutions (not surprisingly) include some non-commutative algebra and logics. We plan to present these developments separately. At present our implemented tools include the following features: adding new elements compatible with the main animation, and (under development) control of facial expressions, of the style and intensity of the motions, etc.

V. Current Implementation

Our Photo-Animation method has been implemented in a number of animation tools. Part of them work online, and they have been incorporated into a site http://photo-animator.weizmann.ac.il.

The site has two lines. One (Live Photo) provides a completely automatic transformation of a photo into a song-based video clip. In particular, it incorporates automatic face detection and face model fitting, and automatic adaptation of the animation template scenarios to the specific photos.

The second line (Video Card - VCard) is based, at present, on an interactive-automatic fitting of enface or profile human model to the photo-character. The resulting virtual actor is automatically cut out and inserted into one of the library template scenarios. Figure 8 shows an example of input photo and a frame of the output clip, created by VCard tool.

Fig.8. An input photo and a frame of the output clip.

VI. Conclusion

In the report we present a method and an animation tool for transformation of a photo into an animated video clip. 3D-2D global scale kinematic models for the patterns of human bodies are presented. These models allow an adaptation to specific parameters of the patterns and provide a high visual quality in reproduction of motions, including relatively large 3D rotations.

Tools and methods for a translation of true 3D motion (like various scenarios of human 3D motion presented in Carnegie Mellon University database) into animation of our models are described.

Tools and methods for real-time automatic fitting of our models to image objects are presented, and high level scenarios and their automatic adaptation to the actual position of the characters in the photo are shortly discussed.

Running examples and on-line Photo-animation tool, which implements the presented technology, can be found following indicated links.

VII. ACKNOWLEDGMENT

The authors would like to thank E. Kagan for his support and numerous inspiring discussions.

REFERENCES

[1] Adobe Flash: http://www.adobe.com/products/flash.html.

[2] Anime Studio: http://anime.smithmicro.com.

[3] A. O. Balan and M. J. Black, "The naked truth: estimating body shape under clothing," In D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part II, LNCS 5303, 2008, pp. 15-29.

[4] CMU graphics lab motion capture database. 2007, http://mocap.cs.cmu.edu/.

[5] E. Cossato and H. P. Graf, "Sample-based synthesis of photo-realistic talking heads," in Proc. 1998 Conf. Computer Animation, pp. 103-110.

[6] M. Dontcheva, G. Yngve, and Z. Popovic, "Layered acting for character animation," ACM Trans. Graphics, vol. 22 no. 3, 2003, pp. 409-416.

[7] H. Fujita and M. Arikava, "Animation of mapped photo collections for storytelling," IEICE Trans. Information and Systems, vol .E91-D, no. 6, 2008, pp. 1681-1692.

[8] D. Haviv. Representation and Processing of Images based on Geometric Models, Ph.D. Thesis, Weizmann Institute of Science, 2010.

[9] D. Haviv and Y. Yomdin, "Novel vectorization method for high resolution images," in Proc. 2012 IEEE 27-th Conv. EEEI (this volume).

[10] A. Hornung, E. Dekkers, and L. Kobbelt, "Character animation from 2D pictures and 3D motion data," ACM Trans. Graph. vol. 26, no. 1, 2007, Article 1.

[11] M. Kim, K. Hyun, J. Kim, and J. Lee, "Synchronized multi-character motion editing," ACM Trans. Graphics, vol. 28, no. 3, 2009, Article 79.

[12] L. Pishchulin, A. Jain, M. Andriluka, T. Thormaehlen, and B. Schiele, "Articulated people detection and pose estimation: reshaping the future," in Proc. 2012 IEEE Conf. CVPR, pp. 1-8.

[13] L. Sigal and M. J. Black, "Guest editorial: state of the art in image- and video-based human pose and motion estimation," Int. J. Computer Vision, vol. 87, no. 1, 2010, pp. 1-3.

[14] L. Sigal, M. Isard, H. Haussecker, and M. J. Black, "Loose-limbed people: estimating human pose and motion using non-parametric belief propagation," Int. J. Computer Vision, vol. 98, no. 1, 2012, pp. 15-48.

[15] M. Thorne, D. Burke, and M. van de Panne, "Motion doodles: an interface for sketching character motion," in Proc. ACM Conf. SIGGRAPH '07, Article 24.