Shadow Theatre: Discovering Human Motion from a Sequence of Silhouettes

Jungdam Won* Jehee Lee†

Seoul National University

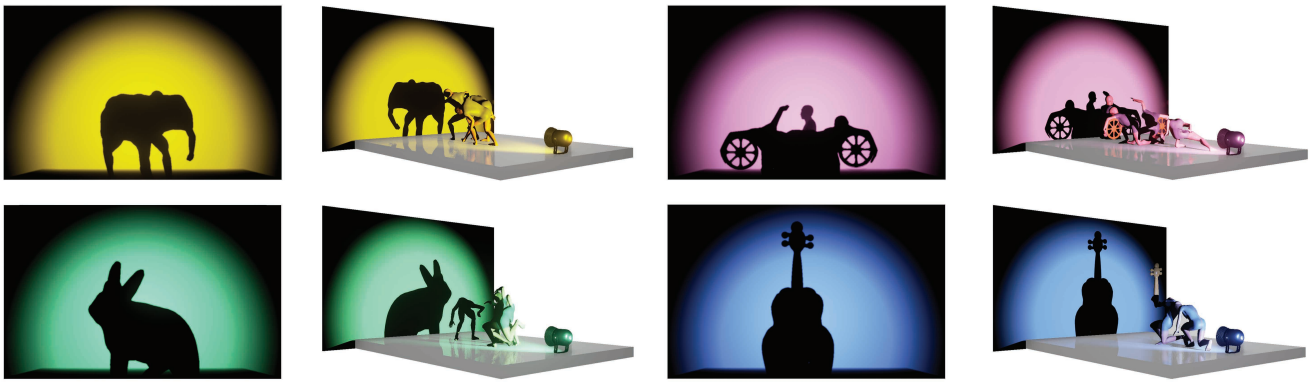


Figure 1: *From a sequence of 2D silhouette images, our shadow theatre system generates animated characters, of which projection on the screen matches well with the target shapes. Our approach is applicable to various shapes such as an elephant (yellow), a rabbit (green), a car (red), and a violin (blue).*

Abstract

Shadow theatre is a genre of performance art in which the actors are only visible as shadows projected on the screen. The goal of this study is to generate animated characters, the shadows of which match a sequence of target silhouettes. This poses several challenges. The motion of multiple characters are carefully coordinated to form a target silhouette on the screen, and each character's pose should be stable, balanced, and plausible. The resulting character animation should be smooth and coherent spatially and temporally. We formulate the problem as nonlinear constrained optimization with objectives, which were designed to generate plausible human motions. Our optimization algorithm was primarily inspired by the heuristic strategies of professional shadow theatre actors. Their know-how was studied and then incorporated into our optimization formulation. We demonstrate the effectiveness of our approach with a variety of target silhouettes and 3D fabrication of the results.

Keywords: Character Animation, Shadow Theatre, Shadow Art, Shadow Play, Multi-Character Coordination, Animation Authoring, Shadows, 2D Silhouettes

Concepts: •**Computing methodologies** → *Motion capture; Motion processing;*

*e-mail:nonaxis@mrl.snu.ac.kr

†e-mail:jehee@mrl.snu.ac.kr

SIGGRAPH '16 Technical Paper,, July 24-28, 2016, Anaheim, CA,
ISBN: 978-1-4503-4279-7/16/07
DOI: http://dx.doi.org/10.1145/2897824.2925869

1 Introduction

Shadow theatre is a genre of performance art that delivers stories in a visual form. What makes it different form other genres is that it utilizes shadows as the only medium of communication. Because audiences can only see the shadows created by the actors on a screen, this stimulates their imaginations regarding what is happening behind the scene. As a result, the audience is immersed in the performance. Although puppets and human hands have been the primary objects used to compose these shadows, several performance teams have created their work using the whole bodies over recent years. Modern lighting technologies have enabled larger stage areas to be exploited for full-body shadow plays, which may raise the level of immersion.

The aim of this study is to reproduce scenes from shadow theatre performances in an automatic manner. Given a sequence of shadow images or silhouettes, we would like to generate the motion of animated 3D characters that cast shadows matching the target. The settings used in state-of-the-art performances, where multiple actors and environments such as props and platforms come on the scene, are also considered. Using the proposed system, the user can explore plausible and creative poses not only for the shapes already used in modern performances but also for new shapes that have not yet been exploited. The system can also be used to model articulated sculptures for the purpose of shadow art.

The basic components that generate shadows during a performance are the light source, the screen, and the actors. If the relative positions and orientations are selected for the components, the shadows on the screen are determined accordingly. The light and screen are usually fixed, whereas the actors move around the stage. Therefore, coordination between the actors and their poses in the spatio-temporal domain are key elements in constructing the scene. However, there are several challenges to be addressed. The first challenge is that certain target shapes require multiple actors to pose collaboratively while coordinating their shadows carefully. The second challenge is the completeness of the resulting shadows, i.e., not only should the outlines of the shadows match the target shape

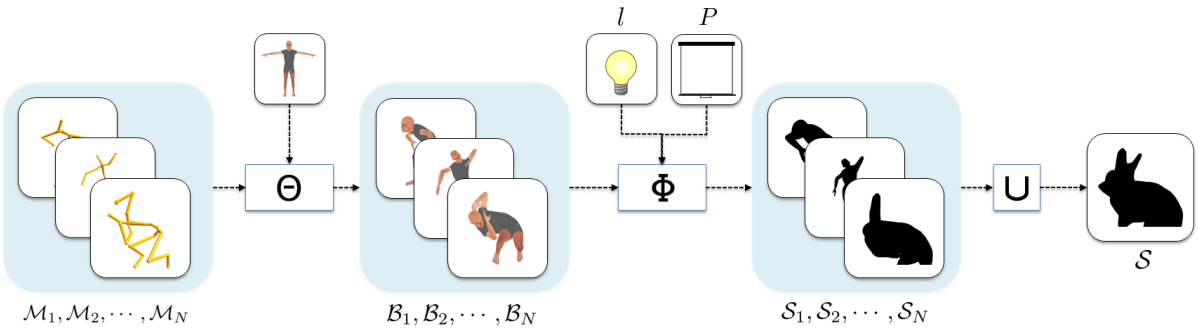


Figure 2: *Shadow generation process. Skeletal motions $\mathcal{M}_1(t), \mathcal{M}_2(t), \cdots, \mathcal{M}_N(t)$, their neutral body meshes and skinning method Θ generate animated body meshes $\mathcal{B}_1(t), \mathcal{B}_2(t), \cdots, \mathcal{B}_N(t)$. Given a stage with a light source l and a screen P, their shadows $\mathcal{S}_1(t), \mathcal{S}_2(t), \cdots, \mathcal{S}_N(t)$ are generated and the final result on the screen is shown as a sum of those shadows.*

but the internal area should also be filled completely. The third challenge is to ensure spatio-temporal coherence of the actors' motions such that smooth target animations can be created. The last challenge is incorporating the nontrivial constraints existing in actual performances (e.g., collisions between actors, physical plausibility of the resultant motions, limited stage area).

The problem is high-dimensional and inherently ill-posed. We formulated it as a nonlinear constrained optimization problem with objectives. The formulation of the objectives was inspired by how professional actors train themselves and perform on stage. We found that they have heuristic strategies, which facilitate our optimization effectively and thus make it feasible. We will demonstrate the effectiveness of our approach with a variety of target silhouettes and animations. The results are further validated by means of 3D fabrication.

2 Related Work

Our work was inspired by the performances of several shadow theatre groups [Attraction ; Fireflies ; Magicplay ; SilhouetteSquad]. A performance typically consists one theme, and it is 2 to 10 minutes long with 20 to 50 scenes. Each scene shows multiple actors, props, and backgrounds. The objects in a scene range from simple ones (e.g., rocks, trees, chairs) formed by a single actor to complex ones (e.g. flowers, elephants, castles) formed by two or more actors. Viewers cannot easily imagine the actors' original poses even for simple objects, and this is the most appealing part of shadow theatre. The origin of shadow theatre goes back to traditional shadow plays, where puppets or human hands were used. Until the 20th century, these shows were performed on a relatively small scale due to the lack of lighting technology. Performance groups including those listed above were formed in the early 2000s in alignment with the advances in technology. These groups are interested in using their entire bodies because it provides a more energetic and delicate feeling, which is difficult to impart using only puppets or hands.

Shadows have long been a key research topic in the computer graphics community because they lend realism and hint at spatial relationships among objects. Many researchers have extensively investigated the rapid and realistic generation of shadows. Crow et al. [1977] proposed the shadow volume technique, which synthesizes shadows by computing cones that enclose a light source and objects. Williams et al. [1978] introduced shadow mapping, which utilizes both pre-rendered depth information obtained from a light source view and a transformation from the light source to the original camera, with this result later used to determine areas of occlusion. These are for local illumination models and the current de facto techniques in real-time graphics applications. For global illumination methods (e.g., ray-tracing, photon mapping, or radiosity), an additional process is not required since its illumination model already reflects the shadow effect [Appel 1968; Whitted 1980; Goral et al. 1984; Jensen 1996].

Shadows for artistic purpose have also been studied. Pellacini et al. [2002] created an interface for placing shadows as user requirements, where the lights or objects are replaced accordingly. Kry et al. [2002] demonstrated an animated hand based shadow puppets by using glove sensors and their real-time skinning technique. Mitra et al. [2009] proposed a tool for sculpting 3D geometries with meaningful shapes when projected onto several non-orthogonal planes, while Bermano et al. [2012] exploited self-shadowing effects to put multiple images into one 3D printed picture, which can be viewed when different light sources are turned on. Mattausch et al. [2013] directly manipulated rendered shadows and applied the edited results in subsequent scene rendering.

The methods for synthesizing collages or mosaics are also relevant to our work from the perspective that the resultant shadow is a collage of shadows made by human bodies. Hausner et al. [2001] created a decorative mosaicking algorithm by placing tiles along user-selected edge features based on a direction field. Collages in non-image domains have also been studied. Gal et al. [2007] modeled expressive 3D characters using primitive 3D models. Kim et al. [2012] synthesized crowd scenes by tiling motion patches that are primitive motion units in the scenes.

There have been many previous approaches which estimated human poses from 2D information (e.g., silhouettes, binary images, sketches). One common approach is to construct a mapping function from 2D information into original 3D poses using human motion databases [Lee et al. 2002; Agarwal and Triggs 2004; Poppe and Poel 2006; Poppe 2007; Ek et al. 2008; Shotton et al. 2013]. Another class of approaches is directly extracting poses by minimizing the visual differences between given 2D inputs and the projection of estimated 3D poses [Sminchisescu and Telea 2002; Davis et al. 2003; Jain et al. 2009; Guan et al. 2009; Wei and Chai 2011; Lin et al. 2012; Ramakrishna et al. 2012; Guay et al. 2013]. Our work is more closely aligned to the second approach, but a key difference is that their input shadows resemble humans, whereas ours do not. Our inputs could be anything other than humans; thus, the estimated poses are often uncommon and acrobatic.

Figure 3: *Experiments with professional actors. We provided target shapes and recorded how they pose to cast a matching shadow. Here, two actors initially shaped an outline of a triangle and another one crouching at the front filled the remaining holes inside.*

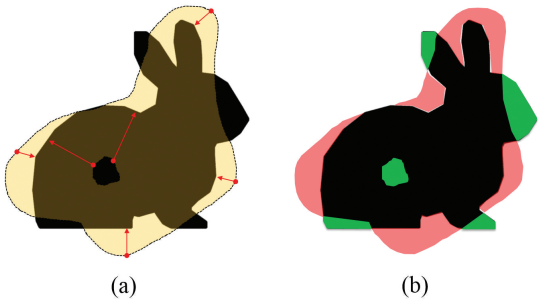


Figure 4: *Contour and coverage differences. (a) The contour-related terms are computed between points sampled on one contour and their closest points on the other contour. (b) The coverage terms minimize the mutual subtraction of the shape areas. $(\mathcal{S} \setminus \mathcal{T})$ and $(\mathcal{T} \setminus \mathcal{S})$ are shown in red and green, respectively.*

3 The Shadow Theatre Problem

A scene in shadow theatre is composed of a set $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_N$ of N actors' shadows on screen plane $P = (\mathbf{n}, d)$, where $\mathbf{n} \in \mathbb{R}^3$ and $d \in \mathbb{R}$ are the normal vector and a constant respectively, of the plane (see Figure 2). Each actor's shadow $\mathcal{S}_n$ is determined by its original body shape $\mathcal{B}_n$ in the 3D space and the projection function $\Phi : \mathcal{B}_n \rightarrow \mathcal{S}_n$, which is defined using a point light source $l \in \mathbb{R}^3$. The body shape $\mathcal{B}_n$ is derived from a 3D skeletal posture $\mathcal{M}_n = (\mathbf{v}_0, \mathbf{q}_0, \mathbf{q}_1, \cdots, \mathbf{q}_M)$ and a mesh skinning function $\Theta : \mathcal{M}_n \rightarrow \mathcal{B}_n$, where $\mathbf{v}_0 \in \mathbb{R}^3$ and $\mathbf{q}_0 \in \mathbb{S}^3$ are the position and orientation of the root joint respectively, $\mathbf{q}_m \in \mathbb{S}^3$ for $m > 0$ is the relative orientation of joint m with respect to its parent joint, and Θ is a linear skinning model. The input to our system is a target shadow $\mathcal{T}$, which is represented as a binary image. The output of the system is a set of N actors' full-body poses $\{\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_N\}$ that projects shadows $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_N$ on the screen matching the target shadow. The problem generalizes easily to take a sequence of target shapes (2D shadow animation) as input and produce a coordinated animation of N actors (3D animation of full-body characters).

Approaches of Professional Actors: Discovering 3D full-body poses from a 2D target shape is an inherently ill-posed problem. Many different sets of poses can generate similar shapes when projected onto a screen. The number of unknown variables to be determined is large (N for actors $\times$ M for joints $\times$ T for time frames); the sum exceeds one thousand, even for a few seconds of performance. Moreover, the process from $\mathcal{M}$ to $\mathcal{S}$ is highly non-linear and non-intuitive because it is a composition of kinematics, skinning, and projection functions. Therefore, naive approaches such as trial and error or exhaustive search are almost impractical. In order to manage these difficulties, we recorded the training session of professional actors to understand how they pose to match the target shapes we provided (see Figure 3). Through the training session, we learned that they have heuristic yet effective strategies, as summarized below.

Strategy 1 *(Principal poses)*. Given an animation of target shapes, they first determine a representative shape at a time frame, for which they pose to cast shadows. The poses that generate the representative shape are called *principal poses*. The full character animation is generated by adapting principal poses gradually over the other frames.

Strategy 2 *(Distinctive features)*. The actors have extensive prior knowledge about which body parts match particular features (e.g., a snout and ears for a rabbit, a trunk and ears for an elephant) in the target shape. For example, they tend to use elbows and knees to match sharp angles and vertebral or gluteal-femoral areas to match round curves.

Strategy 3 *(View direction)*. The actors should be capable of looking at the screen while posing in performances. Since shadow performance requires delicate body control and adjustment continuously, the constraint on viewing direction is crucial.

Strategy 4 *(Balance)*. Balance and stability are also important issues. If several poses are equally plausible for a target shadow, more stable poses are preferred.

Strategy 5 *(Scaling)*. Although the shadow cast by a point light can be scaled easily by moving the actors between the light and the screen, the size of the target shadow matters if the stage area is limited. Actors use a simple rule stating that fewer should join for smaller shapes and more should join for larger shapes.

Strategy 6 *(Minimal motion)*. Once principal poses at the representative frame are determined, the actors adjust their poses to track a sequence of animated target shapes continuously. While doing so, they tend to minimize joint movements and deviation from the principal poses. They typically do not change their ground contact states while tracking a continuous sequence in order to maintain their balance and stability.

4 Discovery of Principal Poses

Given a 2D target shape, we formulate the discovery of principal poses as non-linear optimization of an energy function, which is designed to reflect the heuristic strategies and stage constraints in performances. The principal poses of N characters are computed simultaneously in a single energy optimization.

$$\underset{\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_N}{\operatorname{argmin}} \quad E_{\text{contour}} + E_{\text{coverage}} + E_{\text{coherence}} + E_{\text{visible}} + E_{\text{physics}} + E_{\text{hint}} \tag{1}$$

$$\text{subject to } \mathbf{a_l} \leq \Gamma(\mathcal{M}_n) \leq \mathbf{a_h} \text{ for } n = 1, 2, ..., N.$$

Here, $\{\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_N\}$ denotes a set of full-body poses, Γ is a joint angle measure function. $\mathbf{a_h}$ and $\mathbf{a_l}$ are the upper and lower

bounds for all joints, respectively. We employed axial, conic, and spherical constraints in the unit quaternion space to design the joint measure function [Lee 2000]. The first term E_{contour} penalizes visual differences between the resultant and target shadow contours on the projection screen.

$$E_{\text{contour}} = w_1 \sum_i \|\mathbf{p}_i - \mathbf{p}'_i\|^2 + w_2 \sum_i \|\mathbf{n}_i - \mathbf{n}'_i\|^2 + w_3 \sum_i \|\kappa_i - \kappa'_i\|^2 \quad (2)$$

where $\mathbf{p}_i$ is a point sampled on the contour of the target shape and $\mathbf{p}'_i$ is its closest point on the resultant shadow contour. $\mathbf{n}_i$ and κ_i are the normal and curvature at point $\mathbf{p}_i$, respectively.

The second term E_{coverage} measures the area difference between the resultant and the target shape (see Figure 4).

$$E_{\text{coverage}} = w_4(\text{Area}(\mathcal{T} \setminus \mathcal{S}) + \text{Area}(\mathcal{S} \setminus \mathcal{T})) \quad (3)$$

where $\mathcal{S}$ is the shadow of the characters, $\mathcal{T}$ is the target shape, and $\setminus$ represents the area difference operator on the 2D geometry. This term is a supplement to the contour difference for shape matching, and it also prevents unexpected holes inside the contour.

It is preferred in shadow performances for the shadow of each individual actor to adhere to a coherent sub-section of the target contour, as being responsible for multiple, scattered sections is burdensome. The third term $E_{\text{coherence}}$ encourages the contour to be divided into a small number of pieces such that the shadow of each individual character corresponds to a few continuous pieces,

$$E_{\text{coherence}} = w_5 \sum_i \text{CountIf}(\text{Whose}(\mathbf{p}_i) \neq \text{Whose}(\mathbf{p}_{i+1})) \quad (4)$$

where $\text{Whose}(\mathbf{p}_i)$ provides the index of an actor whose shadow is closest to point $\mathbf{p}_i$. CountIf returns one if the argument is true, and zero otherwise.

The fourth term E_{visible} favors poses with their viewing direction towards the projection screen,

$$E_{\text{visible}} = \begin{cases} w_6 \sum_n \|\mathbf{z} - \mathbf{z}'_n\|^2, & \text{if the cone reaches the screen} \\ \infty, & \text{otherwise} \end{cases} \quad (5)$$

where $\mathbf{z}$ is the center of the screen and $\mathbf{z}'_n$ is the closest point on a viewing cone that approximates the visible volume of the n-th actor (see Figure 5).

The fifth term E_{physics} favors poses that are physically plausible. Each character should be in contact with the floor, balanced, and robust against external perturbations.

$$\begin{aligned} E_{\text{physics}} = w_7 \sum_n \text{DistFloor}(\mathcal{B}_n) + \\ w_8 \sum_n \text{Area}(SP_n) + \\ w_9 \sum_n \text{Margin}(SP_n, COM_n) + \\ w_{10} \sum_n \sum_m \text{PtDepth}(\mathcal{B}_n, \mathcal{B}_m) + \\ w_{11} \sum_n \sum_k \text{PtDepth}(\mathcal{B}_n, \mathcal{E}_k) \end{aligned} \quad (6)$$

where DistFloor computes the minimum distance between the body of an actor and the floor. SP_n and COM_n are the support polygon and center of mass projected on the ground of the n-th actor, respectively. The support margin (Margin), which is the signed distance from COM_n to the boundary of SP_n, measures the robustness of a statically-balanced posture. Additional terms are used to deal with person-to-person and person-to-environment (e.g., floor, screen and props) collisions. $\mathcal{B}_n$ is a skinned body mesh and $\mathcal{E}_k$ is the geometry of an environment object. The penetration depth (PtDepth) measures the degree of interpenetration between objects. Note that PtDepth also measures the self-penetration depth if the same object is given.

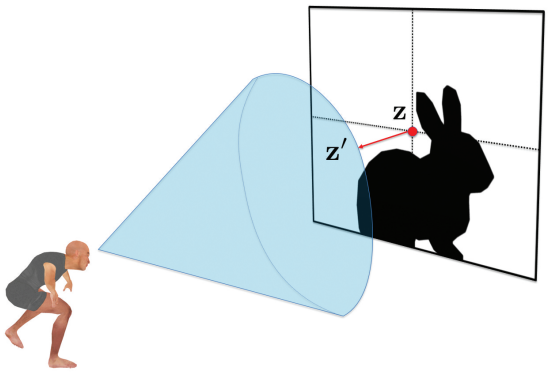


Figure 5: *Visibility: The optimization minimizes the distance from the center of the screen to the viewing cone.*

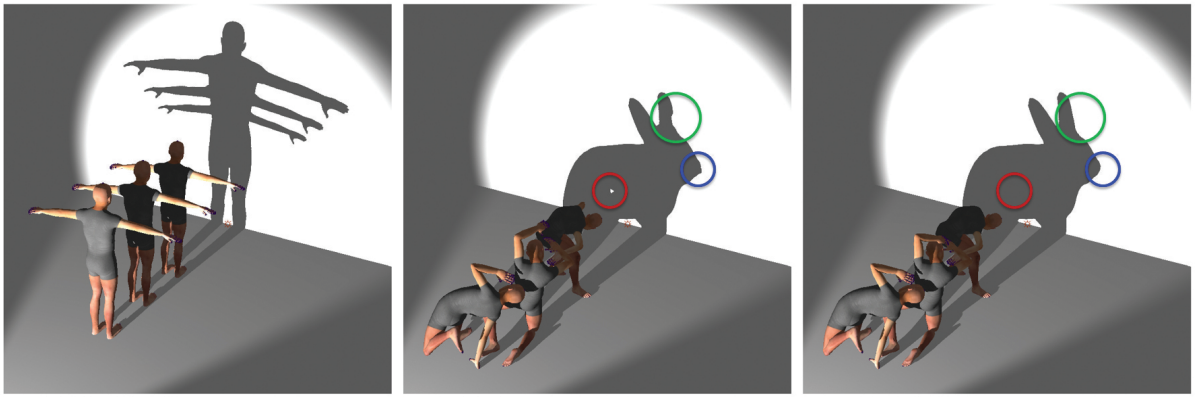
E_{hint} is a term which allows the user directly to interfere with the optimization process. The user can exploit heuristic knowledge with this term, as described in *Strategy 2*,

$$E_{\text{hint}} = w_{12} \sum_i \min(\|\mathbf{h}_i - \mathbf{h}'_{i1}\|^2, ..., \|\mathbf{h}_i - \mathbf{h}'_{ik}\|^2) \quad (7)$$

where $\mathbf{h}_i$ is a hint point on the target contour and $\mathbf{h}'_{ik}$ is a matching point on the character's body. For example, $\mathbf{h}_i$ is an ear tip of a bunny shadow, and we may want to use a character's elbow to match the tip. Then, $\{\mathbf{h}'_{i1}, \cdots, \mathbf{h}'_{ik}\}$ includes the right and left elbows of all characters on the stage. E_{hint} attracts the closest elbow to the hint point.

Optimization Algorithm. The CMA-ES [Hansen and Ostermeier 1996] was adopted to solve our optimization problem. Although CMA-ES was shown to be a powerful nonlinear solver in many previous studies, it often converges to sub-optimal solutions for our problem due to its high-dimensionality and nonlinearity. Sub-optimal solutions can cause visual artifacts, such as holes in shadows and contour mismatches. Our optimization algorithm takes CMA-ES as a basis and adds extra steps to escape from local extrema. If CMA-ES converges to a solution and the residual energy in equation 1 is above a user-specified threshold or visible artifacts are present in the solution, the extra steps are activated to search for better solutions based on two key ideas (see Figure 6).

The first idea is to exploit body parts that are completely occluded by other body parts, other actors, or props. Such occluded body parts do not contribute to forming shadow contours or filling holes can thus be manipulated freely. Starting from individual end-effectors, we check if the body part is completely occluded and move on to its parent link to repeat. In this way, we can identify a chain of completely occluded body parts sequentially. Re-running the CMA-ES algorithm with only the occluded body chain, while leaving the remaining parts fixed, may refine the solution. If there is no such chain, the second idea is to reduce the dimensionality of



(a) (b) (c)

Figure 6: *Refinement of principal poses. (a) All actors initially T-pose in a row. (b) CMA-ES generates poses with artifacts such as holes and silhouette mismatches. The foremost actor's left arm was completely occluded by the other body parts, so does not contribute forming the shadow. (c) The local refinement steps moved the occluded arm to fill in the hole and matched the shadow contour more precisely.*

the problem by selectively including degrees of freedom that may contribute to removing visual artifacts immediately. To do so, the body parts, of which shadows are contiguous to holes or contour mismatches, are selected for the refinement. CMA-ES with reduced degrees of freedom often converges to better solutions. If the refinements are still unsatisfactory, the last resort is to add a new actor to the scene. We add actors one-by-one and repeat the whole process until the resultant principal poses are satisfactory.

5 Animating Principal Poses

In this section, we discuss how to deal with a series of target shadows to generate character animation. Our algorithm follows the strategies of professional actors; we first select a representative frame for which the algorithm from the previous section generates principal poses. The principal poses serve as an initial guess for optimizing the poses at neighboring frames. This optimization process propagates to neighboring frames until a full animation is generated. Without loss of generality, we assume that the representative frame is t_0 and the optimization propagates forward to compute poses at frames t_i for $i > 0$.

Initial Configuration. Providing the optimization solver with a good initial guess is important, particularly when the problem is high-dimensional and nonlinear. We assume that the interiors of the target shapes are triangulated into K pieces consistently (see Algorithm 1). Let $\mathcal{T}^i = \mathcal{T}(t_i)$ be a target shape at frame t_i with its interior triangulated. Let $\mathcal{M}_n^i$ be a pose of the n-th character at frame t_i and $\hat{\mathcal{M}}_n^i$ be a long vector concatenating all joint positions in the reference system. Applying the forward kinematics map to $\mathcal{M}_n^i$ produces $\hat{\mathcal{M}}_n^i$. The light source and the target shape at frame t_i forms a generalized cone $\mathcal{G}^i = \mathcal{G}_1^i \cup \cdots \cup \mathcal{G}_K^i$ of K tetrahedrons, where each tetrahedron is composed of the light source and one of the interior triangles in $\mathcal{T}^i$ (see Figure 7). Then, any joint of a character is positioned in a tetrahedron, and its barycentric coordinates in the tetrahedron specify the joint position in the generalized cone $\mathcal{G}^i$. Let C_n^0 be the barycentric coordinates of all joint positions of principal poses $\hat{\mathcal{M}}_n^0$ with respect to generalized cone $\mathcal{G}^0$ (line 2–3).

For any $\mathcal{T}^i$, we would like to estimate the character's poses roughly with the initial barycentric coordinates C_n^0 with respect to generalized cone $\mathcal{G}^i$ at frame i (line 7). Any joint at frame t_0 has a tetrahedron $\mathcal{G}_k^0$ containing the joint and its barycentric coordinates. This joint is mapped to a new position at frame t_i, which is determined using the barycentric coordinates in $\mathcal{G}_k^i$. Given the joint estimates, determining the poses at a new frame can be formulated as an inverse kinematics problem (line 8). We used a standard technique based on damped Jacobian pseudo inverse and line minimization to solve for inverse kinematics.

Algorithm 1 Adapting principal poses for new target shapes

t_0 : representative frame
$\mathcal{T}^i$: target shape at frame t_i
$\mathcal{G}^i$: generalized cone at frame t_i
$\mathcal{M}_n^i$: the pose (joint configuration) of n-th character
$\hat{\mathcal{M}}_n^i$: joint positions of n-th character

1: **for** $n \leftarrow 1, \cdots, N$ **do**
2: $\quad \hat{\mathcal{M}}_n^0 \leftarrow \text{ForwardKinematics}(\mathcal{M}_n^0)$
3: $\quad C_n^0 \leftarrow \text{BarycentricCoord}(\mathcal{G}^0, \hat{\mathcal{M}}_n^0)$
4: **end for**
5: **for** $i \leftarrow 1, \cdots, T$ **do**
6: $\quad$ **for** $n \leftarrow 1, \cdots, N$ **do**
7: $\quad\quad \hat{\mathcal{M}}_n^i \leftarrow \text{BarycentricCoord}^{-1}(\mathcal{G}^i, C_n^0)$
8: $\quad\quad \mathcal{M}_n^i \leftarrow \text{InverseKinematics}(\hat{\mathcal{M}}_n^i)$
9: $\quad$ **end for**
10: $\quad \mathcal{M}^i \leftarrow \text{Optimize}(\mathcal{T}^i, \{\mathcal{M}_n^i\})$
11: **end for**

Optimization for Motion Generation. Once the poses at frame i are estimated, the next step is to run optimization with those poses as the initial configuration (line 10). The optimization process for motion generation is similar to the one in the previous section except that three additional energy terms (E_{smooth}, E_{regul}, $E_{contact}$) are exploited. E_{smooth} ensures temporal coherence between frames preventing jerky animation,

$$E_{smooth} = w_{13} \sum_n \text{Diff}(\mathcal{M}_n(t-1), \mathcal{M}_n(t)), \tag{8}$$

where the pose difference at two frames measures the discrepancy of the root positions and the joint angles.

$$\text{Diff}(\mathcal{M}(t_1), \mathcal{M}(t_2)) = w^v \|v_0(t_1) - v_0(t_2)\|^2 + \sum_m w_m^q \|q_m(t_1)^{-1} q_m(t_2)\|^2. \tag{9}$$

	Hints		Weights			
	Location	Body parts	w_1	w_2	w_3	w_4
Triangle	n/a	n/a	0.01	0.005	0.005	1000
Rectangle	n/a	n/a	0.01	0.005	0.005	1000
Circle	n/a	n/a	0.01	0.005	0.005	1000
Elephant	Nose tip	Hands	0.015	0.01	0.005	1000
Hat	n/a	n/a	0.01	0.005	0.005	1000
Highheel	Heel center	Knees	0.01	0.01	0.01	1500
Mountain	n/a	n/a	0.01	0.005	0.005	1000
Rabbit	Left ear tip	Elbows	0.01	0.01	0.005	1500
	Right ear tip	Elbows				
	Nose tip	Elbows, Shoulders				
Tshirt	n/a	n/a	0.01	0.01	0.005	1000
Car	Front bumper	Hands	0.01	0.01	0.01	1500
	Rear bumper	Hands				
	Front window tip	Hands				
Airplane	Left wing	Hands	0.02	0.01	0.01	1000
	Right wing	Hands				
	Nose tip	Hands				
Violin	Body center	Hands	0.01	0.005	0.005	1000

Table 1: *Hint specifications and weight values. No hints were provided for Triangle, Rectangle, Circle, Hat, Mountain, T-shirt examples. Weight values of w_1 to w_4 were adjusted, while the other weights were fixed for all examples ($w_5 = 5.0, w_6 = 0.1, w_7 = 3.0, w_8 = 0.001, w_9 = 10.0, w_{10} = 0.04, w_{11} = 0.04, w_{12} = 0.25, w_{13} = 300.0, w_{14} = 40.0$).*

Here, $\|q_m(t_1)^{-1}q_m(t_2)\|$ is the geodesic distance between unit quaternions $q_m(t_1)$ and $q_m(t_2)$. w^v and w^q_m weigh the significance of individual degrees of freedom.

The regularization term E_{regul} prevents excessive deviations from initial principal poses,

$$E_{regul} = w_{14} \sum_n \mathrm{Diff}(\mathcal{M}_n(t_0), \mathcal{M}_n(t)) \tag{10}$$

where t_0 is the index of the representative frame where the principal poses are computed.

The contact term $E_{contact}$ prevents contact changes in character animation based on *Strategy 6* thus, the initial contact states at the representative frame remain unchanged throughout the optimization process.

$$E_{contact} = \sum_{j \in \text{contact}} w_{15}\|c^v_j(t_0) - c^v_j(t)\|^2 + w_{16}\|c^q_j(t_0)^{-1}c^q_j(t)\|^2 \tag{11}$$

where c^v_j and c^q_j are the position and orientation, respectively, of the j-th body part with respect to the reference system, if the body part is in contact with the ground surface.

6 Results

In this section, we demonstrate our shadow generation algorithm using various target silhouettes, their animations, and physical fabrications of the results (Figures 8–12). The target silhouettes include simple geometric shapes such as triangles, rectangles and circles of different sizes (see Figure 8), and complex figurative shapes such as rabbits and elephants (see Figure 9). Hints and weight values for all examples are summarized in Table 1.

Since CMA-ES is based on stochastic sampling, optimization would converge to different solutions with different random seeds

		Expected trials	Time per trial	Total time
Simple Shapes	Triangle	1.25	3.48	4.35
	Rectangle	1.25	3.47	4.34
	Circle	2.5	3.65	9.13
Complex Shapes	Elephant	25	6.69	167.25
	Hat	3.33	5.82	19.38
	Highheel	12.5	6.6	82.5
	Mountain	8.44	6.87	57.98
	Rabbit	33.33	6.98	232.64
	Tshirt	23	7.37	169.51
	Car	12.5	6.51	81.38
	Airplane	2.5	5.93	14.83
	Violin	3.33	5.84	19.45

		Key frames	Time per frame	Total time
Animations	Rabbit	16	4.5	72
	Elephant	34	3.66	124.44
	Car	12	3.86	46.32

Table 2: *Runtime performance. The computation time is measured in minutes.*

even if the algorithm begins with the same initial configuration. Therefore, we select the best result among multiple optimization trials. The population size of CMA-ES is $\lambda = 40$. The optimization at each trial converges within one thousand iterations for simple shapes and within two thousand iterations for complex shapes. Only a few hundred iterations are sufficient for animation generation because there is strong frame-to-frame coherence.

The algorithm runs on a desktop PC equipped with an Intel Core i7 4790K (4 cores, 4.0GHz) and NVIDIA GeForce GTX 480. Table 2 presents the performance statistics. The table shows expected trials until we get one satisfactory result. If the expected trials is 3, one out of three results is satisfactory. The notion of satisfaction is admittedly subjective and dependent on our artistic sense. Our criteria include two measures. First, the result should be recognizable as was intended. For example, an elephant's silhouette should be recognizable as an elephant for any viewer. Second, the result should reproduce the key features of the target silhouette. To discover the principal poses, our algorithm can find plausible solutions for most of our examples within a few trials and 5 to 20 minutes of computation except for several challenging examples. The rabbit in Figure 9 is the most difficult, requiring dozens of trials until a satisfactory result is obtained and 4 hours of computation time. Although the time complexity is basically proportional to the number of characters and the number of key features in the target shape, there are other factors that may affect the computation time significantly. For example, the inverted triangle has a wide horizontal span at the top and narrow support at the bottom, which makes it difficult to find balanced stable poses. The silhouette of the t-shirt in Figure 9 is simpler than those of the elephant and the car, but the t-shirt requires more computation time since it has narrow support, which makes the characters pose acrobatically. To generate animations, only one trial is enough for each key frame because we start the adaptation process with plausible initial poses. It takes 50 to 125 minutes to generate entire animations, and computation time for a single optimization is reduced as compared to discovering principal poses due to its fast convergence.

Implementation Details. The implementation of the system includes many technical components and acceleration techniques. An open source game engine [OGRE] was used to implement charac-

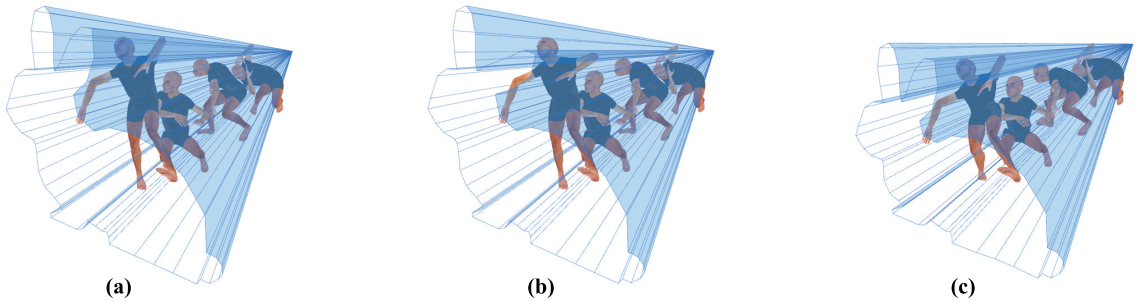


Figure 7: *Motion generation. (a) The principal poses at the representative frame. The rays from the light source form a generalized cone with its base matching the target silhouette. (b–c) adapting the principle poses to similar target shapes. The poses change smoothly and coherently across frames.*

ter skinning and shadow rendering. The built-in shader functionality of the engine was modified to color-code the pixels in the shadows by character identifiers. In order to compute the penetration depth efficiently, we approximate the 3D character's body volume using spheres. The $E_{contact}$ term requires CMA-ES to solve inverse kinematics as a part of a big stochastic optimization. We decouple inverse kinematics from the optimization and run a standard inverse kinematics solver for every CMA-ES sample to manage the contact term separately.

Animation. Beginning with a 2D silhouette, we used a shape manipulation technique [Igarashi et al. 2005] to generate 2D shadow animation (see Figure 12 and the supplemental video). For the rabbit and elephant examples, the target animations were designed to have large movements of their perceptual features. The rabbit moved its ears back and forth while lowering its head. The elephant raised its trunk as if trumpeting. For the car example, we choreographed a scene with a person driving a car, which is jerking up and down.

3D Fabrication. We 3D-printed principle poses using an *Object 24* printer (see Figure 11). The mesh of a resultant pose requires post-processing steps to be fed into the 3D printer. We remove the self-collision of the mesh, which is an artifact of linear skinning deformation. We also reinforce fragile support between the character's feet and the ground. Two lighting devices were tested for stage installation. The first device was a *Cree* LED light module, which is small yet bright to imitate an ideal point light source. It can also be used as a spot light if a convex lens is installed. The second device is a portable projector *Optoma ML750*, which emulates an actual stage setting because professional actors use a large-scale projector as their light source in stage performance. The fabrication results consistently generated shadows similar to the target shapes with both lighting devices.

7 Conclusions and Future Work

Shadow theatre is a large optimization problem, consisting of many non-linear components such as kinematics, kinetics, skinning, perspective projection, and multi-character coordination. The biggest challenge in our work is to make the optimization converge to a plausible solution with reasonable computing resources. The key to success was the design of the objective functions, which compete with each other and thus keep the optimization process balance on plausible sub-manifolds in the vast search spaces. The heuristic strategies learned from professional actors also narrowed down the search space substantially and made the results human-like. Selective refinement of DOFs effectively localized the large optimization to reduce it into a series of smaller optimization tasks.

There are a number of limitations in this study. Artifacts of linear skinning and inaccurate body modeling affect shadow generation particularly when the characters pose acrobatically. More realistic body models based on either biomechanical or data-driven modeling would improve the quality of the results [Anguelov et al. 2005; Lee et al. 2009; Loper et al. 2015]. The computational bottleneck is the rendering of shadow images by the rendering engine. The current high-performance graphics hardware on typical desktop computers can render hundreds of images per second, while our optimization algorithm require 40,000 to 80,000 shadow images per trial. Another bottleneck is set operations between shadow images for the evaluation of $E_{coverage}$ in Equation (3). There is room for performance improvement since GPU implementation of pixel-by-pixel image operations would achieve significant speedups.

We can think of many directions for future research. The problem setting can be generalized to exploit multiple light sources and multiple screens. Multiple light sources would generate shades of gradation at the penumbra, providing new means of expression. With multiple screens, we expect different images to be projected onto the screens from a single character pose. It is probably daunting for human actors to pose while having to consider the casting of shadows in multiple directions. Computer graphics technology can be a tool for the choreography of next-generation performances. We can also think of other possibilities such as curved screens and non-point light sources. In the future, a computer graphics system similar to ours can serve as a test bed for a variety of stage installations and performance scenarios.

Acknowledgements

We thank all the reviewers for their comments, and the participants in our experiments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.2011-0018340).

References

AGARWAL, A., AND TRIGGS, B. 2004. 3d human pose from silhouettes by relevance vector regression. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, II–882–II–888 Vol.2.

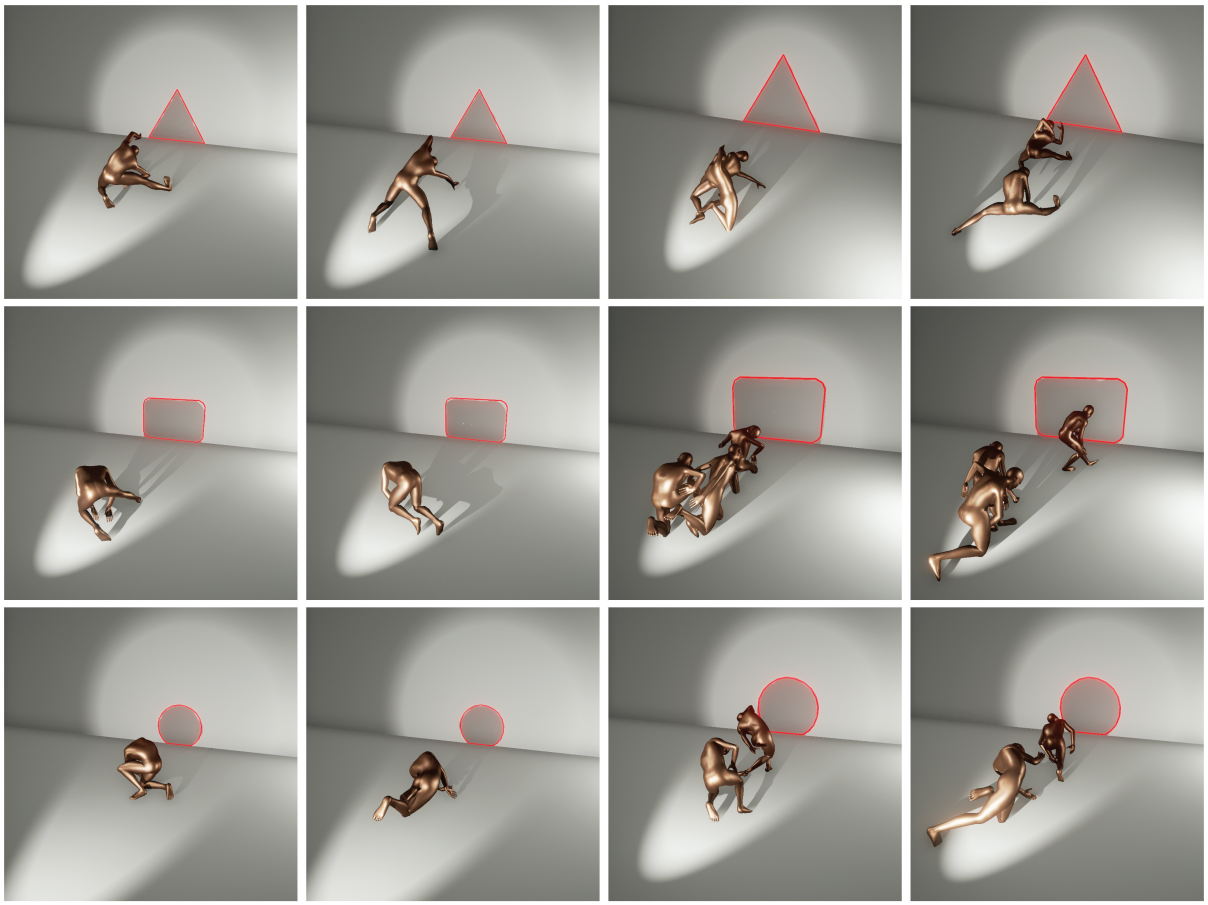


Figure 8: *Principal poses for simple shapes. (1st column) principal poses for target shapes shown in red line. (2nd column) Alternative solutions at another local minima. (3rd and 4th columns) The target shapes are scaled by a factor of 1.5 and thus require more actors to join in.*

ANGUELOV, D., SRINIVASAN, P., KOLLER, D., THRUN, S., RODGERS, J., AND DAVIS, J. 2005. Scape: Shape completion and animation of people. *ACM Transactions on Graphics*, 408–416.

APPEL, A. 1968. Some techniques for shading machine renderings of solids. In *Proceedings of the April 30–May 2, 1968, Spring Joint Computer Conference*, AFIPS '68 (Spring), 37–45.

ATTRACTION. Attraction official website. http://www.attraction.hu/.

BERMANO, A., BARAN, I., ALEXA, M., AND MATUSK, W. 2012. Shadowpix: Multiple images from self shadowing. *Comp. Graph. Forum*, 593–602.

CROW, F. C. 1977. Shadow algorithms for computer graphics. In *Proceedings of the 4th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '77, 242–248.

DAVIS, J., AGRAWALA, M., CHUANG, E., POPOVIĆ, Z., AND SALESIN, D. 2003. A sketching interface for articulated figure animation. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '03, 320–328.

EK, C., TORR, P., AND LAWRENCE, N. 2008. Gaussian process latent variable models for human pose estimation. In *Machine Learning for Multimodal Interaction*, vol. 4892. 132–143.

FIREFLIES. Fireflies official website. http://fireflies.com.ua/.

GAL, R., SORKINE, O., POPA, T., SHEFFER, A., AND COHEN-OR, D. 2007. 3d collage: Expressive non-realistic modeling. In *Proceedings of the 5th International Symposium on Non-photorealistic Animation and Rendering*, NPAR '07, 7–14.

GORAL, C. M., TORRANCE, K. E., GREENBERG, D. P., AND BATTAILE, B. 1984. Modeling the interaction of light between diffuse surfaces. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '84, 213–222.

GUAN, P., WEISS, A., BALAN, A., AND BLACK, M. J. 2009. Estimating human shape and pose from a single image. In *Int. Conf. on Computer Vision, ICCV*, 1381–1388.

GUAY, M., CANI, M.-P., AND RONFARD, R. 2013. The line of action: An intuitive interface for expressive character posing. *ACM Transactions on Graphics 32*, 205:1–205:8.

HANSEN, N., AND OSTERMEIER, A. 1996. Adapting arbitrary normal mutation distributions in evolution strategies: the covari-

ance matrix adaptation. In *Proceedings of IEEE International Conference on Evolutionary Computation*, 312 –317.

HAUSNER, A. 2001. Simulating decorative mosaics. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, 573–580.

IGARASHI, T., MOSCOVICH, T., AND HUGHES, J. F. 2005. As-rigid-as-possible shape manipulation. SIGGRAPH '05, 1134–1141.

JAIN, E., SHEIKH, Y., AND HODGINS, J. 2009. Leveraging the talent of hand animators to create three-dimensional animation. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '09, 93–102.

JENSEN, H. W. 1996. Global illumination using photon maps. In *Proceedings of the Eurographics Workshop on Rendering Techniques '96*, 21–30.

KIM, J., AND PELLACINI, F. 2002. Jigsaw image mosaics. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '02, 657–664.

KIM, M., HWANG, Y., HYUN, K., AND LEE, J. 2012. Tiling motion patches. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '12, 117–126.

KRY, P. G., JAMES, D. L., AND PAI, D. K. 2002. Eigenskin: Real time large deformation character skinning in hardware. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '02, 153–159.

LEE, J., CHAI, J., REITSMA, P. S. A., HODGINS, J. K., AND POLLARD, N. S. 2002. Interactive control of avatars animated with human motion data. *ACM Transactions on Graphics (SIGGRAPH 2002) 21*, 3, 491–500.

LEE, S.-H., SIFAKIS, E., AND TERZOPOULOS, D. 2009. Comprehensive biomechanical modeling and simulation of the upper body. *ACM Transactions on Graphics 28*, 4.

LEE, J. 2000. *A Hierarchical Approach to Motion Analysis and Synthesis for Articulated Figures*. PhD thesis, Department of Computer Science, Korea Advanced Institute of Science and Technology.

LIN, J., IGARASHI, T., MITANI, J., LIAO, M., AND HE, Y. 2012. A sketching interface for sitting pose design in the virtual environment. *IEEE Transactions on Visualization and Computer Graphics 18*.

LOPER, M., MAHMOOD, N., ROMERO, J., PONS-MOLL, G., AND BLACK, M. J. 2015. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics (SIGGRAPH Asia 2015) 34*, 6.

MAGICPLAY. Magicplay official website. http://www.magicplay.co.kr/.

MATTAUSCH, O., IGARASHI, T., AND WIMMER, M. 2013. Freeform shadow boundary editing. 175–184.

MITRA, N. J., AND PAULY, M. 2009. Shadow art. *ACM Transactions on Graphics 28*, 5.

OGRE. Object-oriented graphics rendering engine. http://www.ogre3d.org/.

PELLACINI, F., TOLE, P., AND GREENBERG, D. P. 2002. A user interface for interactive cinematic shadow design. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '02, 563–566.

POPPE, R., AND POEL, M. 2006. Comparison of silhouette shape descriptors for example-based human pose recovery. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, 541–546.

POPPE, R. 2007. Evaluating example-based pose estimation: Experiments on the humaneva sets. In *CVPR 2nd Workshop on Evaluation of Articulated Human Motion and Pose Estimation (EHuM2),*.

RAMAKRISHNA, V., KANADE, T., AND SHEIKH, Y. 2012. Reconstructing 3d human pose from 2d image landmarks. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part IV*, 573–586.

SHOTTON, J., SHARP, T., KIPMAN, A., FITZGIBBON, A., FINOCCHIO, M., BLAKE, A., COOK, M., AND MOORE, R. 2013. Real-time human pose recognition in parts from single depth images. *Commun. ACM 56*, 116–124.

SILHOUETTESQUAD. Silhouette Squad official website. https://www.talentcult.com/SilhouetteSquad/.

SMINCHISESCU, C., AND TELEA, A. 2002. Human pose estimation from silhouettes, a consistent approach using distance level sets. In *In WSCG International Conference on Computer Graphics, Visualization and Computer Vision*.

SORKINE, O., AND ALEXA, M. 2007. As-rigid-as-possible surface modeling. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing*, SGP '07, 109–116.

WEI, X., AND CHAI, J. 2011. Intuitive interactive human-character posing with millions of example poses. *IEEE Comput. Graph. Appl. 31*, 78–88.

WHITTED, T. 1980. An improved illumination model for shaded display. *Commun. ACM 23*, 343–349.

WILLIAMS, L. 1978. Casting curved shadows on curved surfaces. In *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '78, 270–274.

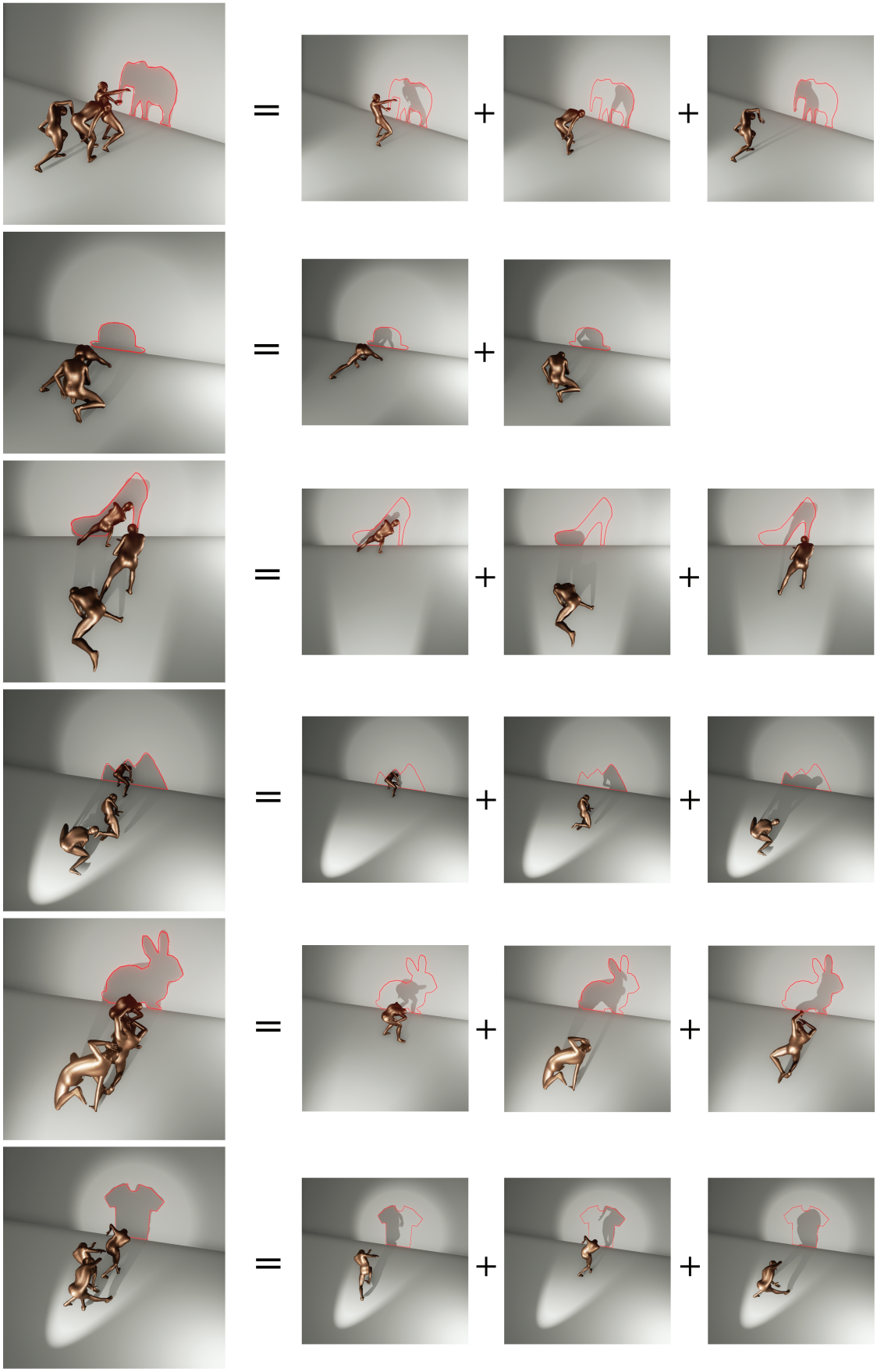


Figure 9: *Principal poses for complex shapes. In each row, the leftmost image shows the optimized poses while the images on the right show the shadows of individual actors.*

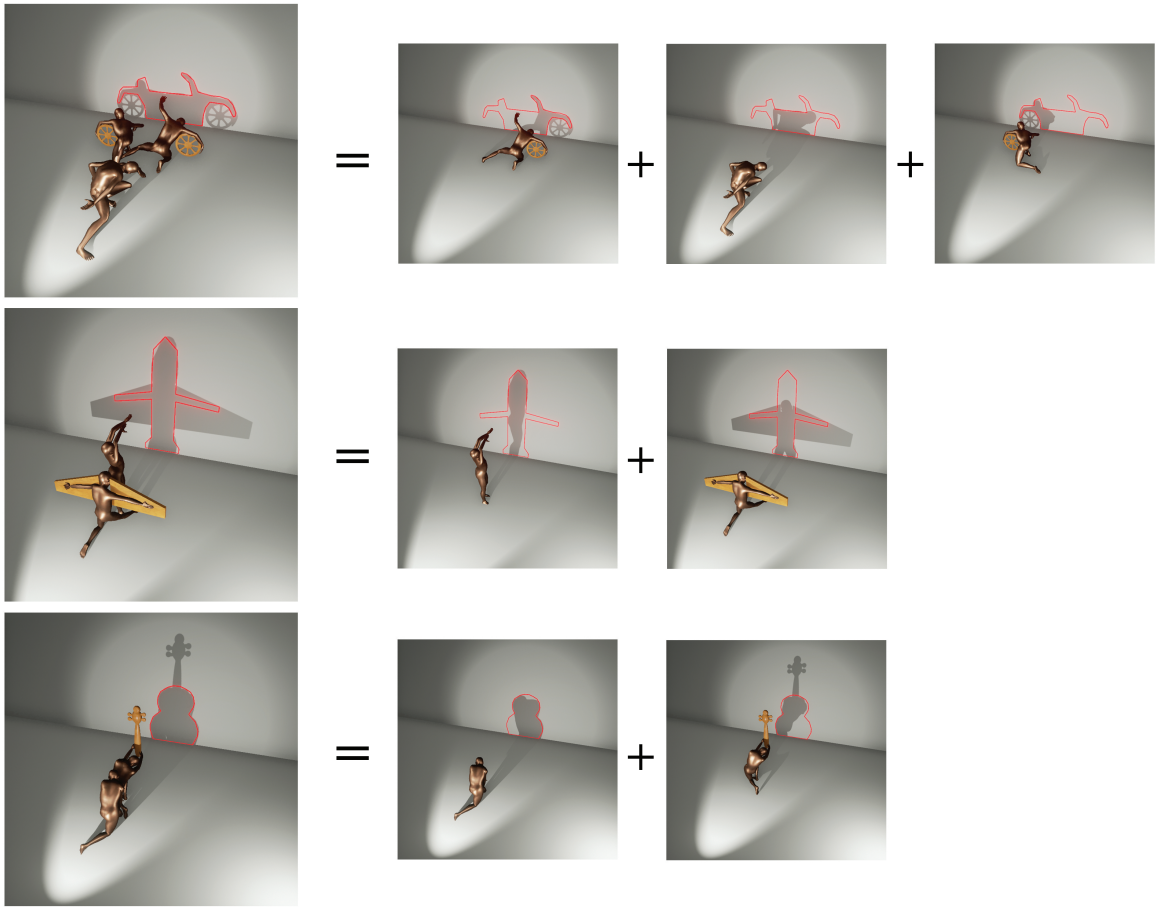


Figure 10: *Posing with props. (Top) Car wheels. (Middle) Airplane wings. (Bottom) Violin neck.*

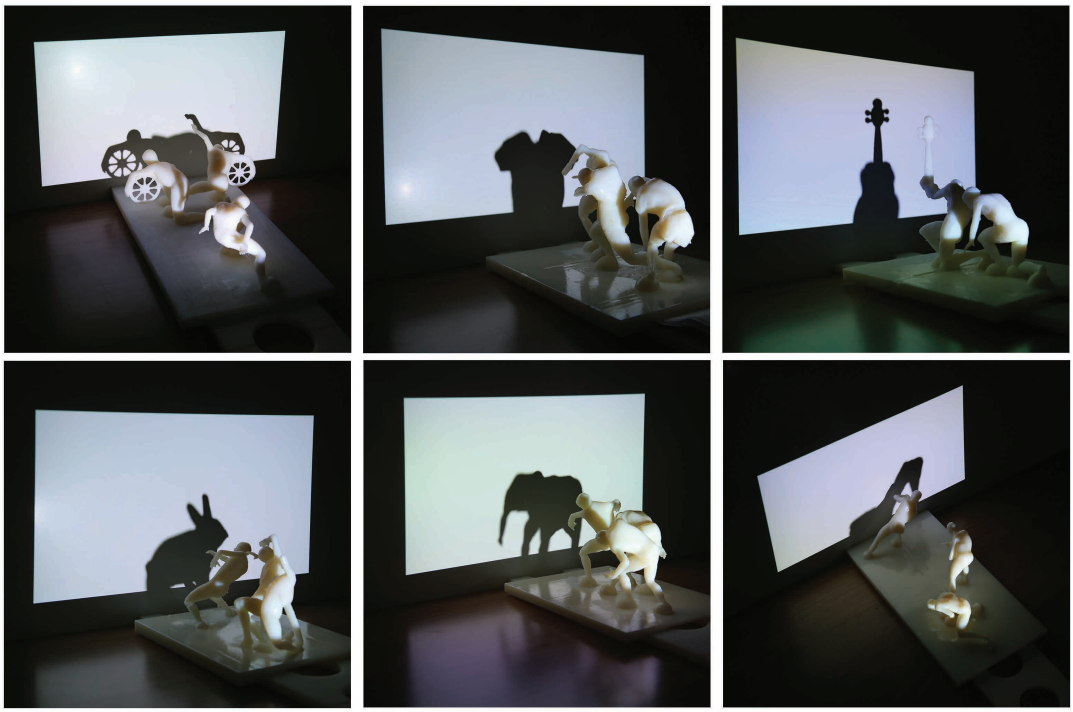


Figure 11: *3D Fabrication results.*

Figure 12: *Shadow animation. (a) Rabbit. (b) Elephant. (c) Car with a driver.*