Motion analysis in video: dolls, dynamic cues and Modern Art

J. P. Collomosse and P. M. Hall

Department of Computer Science, University of Bath, Bath, U.K. BA2 7AY
{jpc|pmh} @ cs.bath.ac.uk

Abstract

This paper addresses the problem of synthesising animations from video clips; in particular emphasising the motion of tracked objects. We introduce "dynamic cues" as a class of motion emphasis cue, encompassing traditional animation techniques such as anticipation and exaggeration. We present methods for automatically synthesising such cues within video premised upon the recovery of articulated figures, and the subsequent manipulation of the recovered pose trajectories. Additionally, we apply our motion emphasis framework to emulate artwork in the Futurist style, popularised by Duchamp.

1. Introduction

Processing real-world video sequences into animation is a challenge that until recently the non-photorealistic rendering (NPR) literature has been almost silent about. There are two main problem: (1) generating stable visual stylisations over the video (for example, painterly effects) ; (2) generating *motion emphasis cues* used by traditional animators. Early attempts to solve the first problem suffered from a distracting flickering [Lit97, HP00] that more recent approaches suppress [CRH05, WXSC04]. A limited range of motion emphasis effects have been produced from three dimensional computer graphics models [CPIS02, BH00], by motion capturing cartoons [BLCD02], or interactively from drawings [SPR*94] and video [AHSS04]; see [Col04] for a wider review. Of greatest relevance to this paper is work addressing the production of both *augmentation cues* and *deformation cues* in real video [CRH03]. The unique contribution of this paper is to extend the analytic framework required for augmentation and deformation cues so that *dynamic cues* can be automatically produced. Furthermore the *Futurist* school of painting, typified by Duchamp, can be emulated; this too is a unique contribution to NPR.

Traditional animators emphasise motion with a variety of cues that are familiar to anyone who has watched animations. Streak-lines depicting the paths of objects, and ghosting effects that echo trailing edges, are both examples of what we call *augmentation cues*: the animation is visually
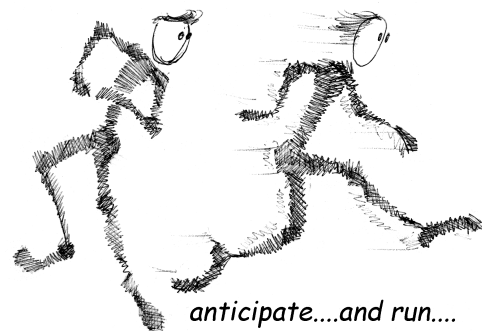

Figure 1: *Anticipation is a common dynamic cue; ghosting and streak-lines are also shown, as is some deformation.*

augmented with marks of some kind. Animated objects may stretch as they accelerate, squash as they slow down, or bend to show drag or inertia — we call these *deformation cues*. Furthermore objects may "anticipate" movement by a slight prior movement backwards, or move in a characteristic way that exaggerates ordinary motion. These latter cues we call *dynamic cues*. Examples of these cues are illustrated in Figure 1. A deeper understanding of the differences between them relies on a definition of pose trajectory, as we now explain.

At any given instant in time an object has a particular *pose*, typically specified by a vector of numbers (for exam-

ple, inter-joint orientations and world position). As this pose vector changes in time we obtain a *pose trajectory*. Augmentation cues and deformation cues are rendered as a function of pose trajectory. Dynamic cues differ because they alter the pose trajectory. This makes rendering dynamic cues very difficult because both the pose and timing of the object may change: poor rendering could leave "gaps" in the video, for example. Furthermore generating dynamic cues is no easy task: a cartoon character can "wind up to run" in a way that is unique to them. The essential simplicities that bind the set of dynamic cues are very difficult to find.

Our purpose here is to provide an initial in-road into an understanding of dynamic cues. To this end we show how to generate and analyse a pose trajectory to produce:

- simple anticipation effects;
- simple motion exaggeration effects;
- novel stills, similar to those of the Futurists, such as Duchamp.

Our broad approach is to track polygons fitted around rigid objects so as to estimate their pose trajectory. This is analysed to construct a hierarchical articulated figure of rigid parts, with its pose trajectory (Section 2). The dynamic cues we produce from this (Section 3) integrate fully with our early published framework for synthesising augmentation and deformation cues [CRH03]. Further, all motion emphasis cues integrate with our stable video stylisation technique [CRH05]. Therefore, the contribution of this paper completes our work in the automated production of animations from real-world video, see [Col04] for a full description of our *Video Paintbox*.

2. Building a doll

Our problem is to recover the motion of a articulated figure — a *doll* — from monocular video. The doll is to be built from rigid parts and have a hierarchical structure. The hierarchy is a tree in which each part corresponds to a tree node. Two nodes are linked in the tree if they are physically connected by a pivot.

Humans are an important class of articulated figures, and the recovery of human motion from video sequences is a well researched problem, see Hicks for a review [Hic03]. Briefly, most techniques use a constraint in the form of many cameras or a prior model of human motion, neither option is open to us for we have one camera and cannot guarantee that a human is the articulated figure. The constraint we use is that the object moves in a plane (more formally: the motion vectors can be sufficiently well represented by a two-dimensional vector space).

The underlying idea is to consider pairs of rigid parts and observe the motion of one relative to the other. This allows us to estimate the centre of rotation, if it exists, at an instant in time. By holding fixed first one object and then the other

we estimate two centres of rotation. If these are sufficiently close and both lie within the intersection of the polygons associated with the rigid parts, then we decide that the two objects are pivoted and select the rotation centre computed when the parent was held still as the pivot point. The root is arbitrarily assigned, its parent is the world frame. A depth ordering between the parts of the figure is assigned using occlusion information available from the video, useful when later compositing features. The tracking and depth recovery processes are beyond the scope of this paper and the reader is referred elsewhere for details [Col04].

Our focus here is to recover the pose trajectory, $\mathbf{p}(t)$ of an articulated object:

$$\mathbf{p}(t) = \begin{bmatrix} \mathbf{c}(t) \\ \theta_1(t) \\ \cdots \\ \theta_n(t) \end{bmatrix} \quad (1)$$

where $\mathbf{c}(t)$ is the location of some identifiable point on the object's root node, and the $\theta_i(t)$ specify the orientation of each branch node relative to its parent; $\theta_1(t)$ orients the whole articulated object using $\mathbf{c}(t)$ as a pivot.

We begin by tracking points on polygons. The *state* of a particle (point) at any time instant, t, is a vector comprising position, $\mathbf{x}(t)$, velocity $\mathbf{v}(t)$ and acceleration $\mathbf{a}(t)$.

$$\mathbf{s}(t) = \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{v}(t) \\ \mathbf{a}(t) \end{bmatrix} \quad (2)$$

Each state is a particle in *state space*. This state is used by the Kalman filter to track objects in video. The reader is referred elsewhere for details of tracking [Kal60].

Given the state of particles on a rigid body (polygon) it is easy to estimate the translation and rotation of the body. At some time t let $\mathbf{x}_i$ be the ith identifiable point of a rigid body. Given three such points these transform, under an instantaneous rotation $\mathbf{R}$ and translate under an instantaneous displacement $\mathbf{u}$. In homogeneous coordinates:

$$\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{u} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x} \\ 1 & 1 & 1 \end{bmatrix} \quad (3)$$

Each matrix is (3×3) so the unknown transform is easy to compute

$$\begin{bmatrix} \mathbf{R} & \mathbf{u} \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 \\ 1 & 1 & 1 \end{bmatrix} \quad (4)$$

Hence we can compute instantaneous changes in orientation and location. It is a simple matter to integrate these to obtain a change relative to the starting orientation to acquire $[\mathbf{c}(t), \theta(t)]^T$, relative to the starting position.

We next consider whether a given pair of rigid objects are pivoted. Given two rigid objects, A and B, we assume the pose trajectory for each of them, $\mathbf{p}_A(t)$ and $\mathbf{p}_B(t)$. Consequently the motion of B relative to A is easy to estimate,
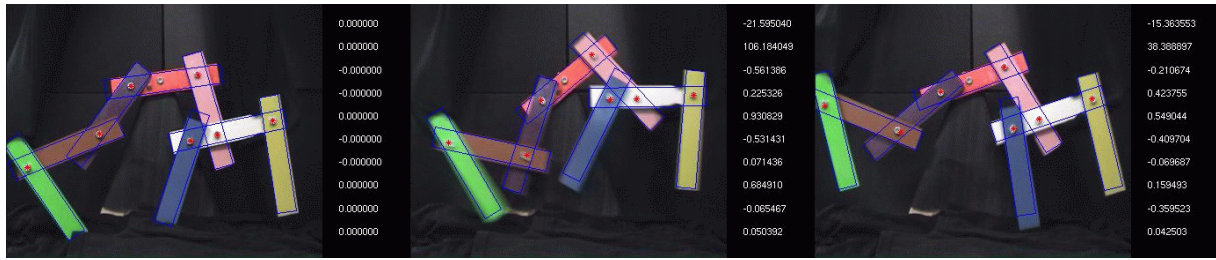
Figure 2: *An articulated contraption used for experiments. The full pose vector at various instants in time is shown alongside.*

being characterised completely by the difference in pose trajectories $\mathbf{p}_B(t) - \mathbf{p}_A(t)$. Therefore we can observe the movement of B in the reference frame of A, which reduces the problem of finding a mutual pivot to one of finding a fixed point about which B rotates (if it rotates at all).

Let $\mathbf{x}_i$ be a point on B, measured in the fixed reference frame of A. Suppose B rotates about the fixed point $\mathbf{f}$, relative to A. If motion is uniform, then after a short time interval dt this point appears at $\mathbf{y}_i$

$$\mathbf{y}_i = \mathbf{R}(\mathbf{x}_i - \mathbf{f}) + \mathbf{f} \tag{5}$$

The problem is to estimate $\mathbf{f}$ given a sufficient number of $\mathbf{x}_i$ and $\mathbf{y}_i$. This problem differs Equation 3 because there rotation about the origin was sufficient, and we computed a translation too; here we seek rotation about an unknown point. We will later discuss the relationship between these two problems in greater depth. The important principle here is $\mathbf{f}$ is a singularity of the transform, therefore we cannot invert the system of equations.

We proceed by solving a system of homogeneous linear equations. Writing x_j for the jth element of some point $\mathbf{x}$, at time t and y_j for the corresponding element at time $t + dt$. Equation 5 becomes

$$y_1 = r_{11}x_1 - r_{12}x_2 + u_1 \tag{6}$$
$$y_2 = r_{21}x_1 - r_{22}x_2 + u_2 \tag{7}$$

in which

$$\mathbf{u} = (\mathbf{I} - \mathbf{R})\mathbf{f} \tag{8}$$

We can now write

$$\begin{bmatrix} x_1 & -x_2 & 0 & 0 & 1 & 0 & -y_1 \\ 0 & 0 & x_1 & -x_2 & 0 & 1 & -y_2 \end{bmatrix} \begin{bmatrix} r_{11} \\ r_{12} \\ r_{21} \\ r_{22} \\ u_1 \\ u_2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{9}$$

It is easy to extend the left-most matrix because each identifiable point in B provides two rows, yielding a design matrix $\mathbf{M}$. The smallest right-singular vector of $\mathbf{M}$ is a suitable solution in the least squared sense. This is normalised so that its seventh element is unity and in this way we obtain the

rotation matrix elements r_{kl} and a displacement $\mathbf{u}$. The pivot $\mathbf{f}$ is obtained from Equation 8 as

$$\mathbf{f} = (\mathbf{I} - \mathbf{R})^{-1}\mathbf{u} \tag{10}$$

Because this estimate of $\mathbf{f}$ is obtained using all identifiable points of B it tends to be robust to measurement error. If there is no rotation, then $\mathbf{R} = \mathbf{I}$, indicating there is no pivot. We decide that B has a pivot relative to A only if a pivot $\mathbf{f}$ exists that lies within the intersection of A and B.

To further improve robustness we reverse the roles of A and B, recomputing the pivot point. Furthermore, we compute the pivot for all time instants t, each over a fixed interval dt. We insist that the pivot remains within the intersection of A and B over all time. Figure 2 illustrates the fact that we can recover complex articulated structures in this way.

We now return to the relationship between Equations 3 and 5. The first of these computes rotation about the origin and an accompanying displacement, the second computes rotation about an unknown fixed pivot. We claim *it is not possible to simultaneously compute a rotation, a pivot and a displacement*. As proof we consider the point $\mathbf{x}$ rotating about the origin with constant angular velocity ω. The tangential velocity of this point is $\dot{\mathbf{x}} = \omega(\mathbf{x} \otimes \mathbf{n})$, where $\mathbf{n}$ is a normal to the plane of rotation and $\otimes$ is vector cross product (it is not necessary for this to obey the right-hand screw rule). Now suppose that $\mathbf{x}$ not only rotates about the origin but translates too, with a linear velocity $\dot{\mathbf{u}}$. The governing equation now is $\dot{\mathbf{x}} = \omega(\mathbf{x} \otimes \mathbf{n}) + \dot{\mathbf{u}}$. Since $\dot{\mathbf{u}}$ is a constant we can always write it in the form $\dot{\mathbf{u}} = \omega(\mathbf{d} \otimes \mathbf{n})$, and therefore obtain $\dot{\mathbf{x}} = \omega(\mathbf{x} \otimes \mathbf{n}) + \omega(\mathbf{d} \otimes \mathbf{n})$. Appealing to the fact that addition distributes over the cross product operator we obtain $\dot{\mathbf{x}} = \omega((\mathbf{x} + \mathbf{d}) \otimes \mathbf{n})$ from which we conclude that effective centre of rotation has been shifted as a consequence of the displacement, in a direction perpendicular to it. This result is analogous to the phenomenon observed in a gyroscope, which when suffering a force in the plane of its rotation moves, in the plane, in a direction orthogonal to the applied force. Here it shows that if we choose an arbitrary pivot we can always determine a compensating displacement, and vice-versa. Therefore we cannot unambiguously estimate both at once; this is an in-principle restriction.

Given a pose trajectory for each rigid body, and a pivot for each pair of linked rigid bodies, it is a matter of book-

keeping to assemble a hierarchical articulated figure, complete with a full pose trajectory of the form in Equation 1; we have automatically assembled a doll from video data.

3. Dynamic cues and Modern Art

Given a recovered doll, we can produce not only dynamic cues as seen in traditional animations, but also emulate the Futurist style of Modern Art. So far as we are aware, both represent unique contributions.

As mentioned the general form of dynamic cues is to map one pose trajectory into another:

$$\mathbf{p}'(t) = \mathcal{F}[\mathbf{p}(t)] \qquad (11)$$

The new pose trajectory is used to govern all other cues, so that objects can be augmented and deformed. Again as mentioned, a full understanding of dynamic cues eludes us at the present time, but we can make some progress by considering two important classes of dynamic cue: anticipation and motion exaggeration. We consider each in turn, followed by a discussion on emulating Futurist art.

3.1. Anticipation

Anticipation is used by animators to make the viewer aware that a change in motion is imminent. This visual cue is highly complex, especially when applied to human gait, changes in facial expression, or the motion of any other complex body.

We interpret anticipation as a change in the pose trajectory as a result of a change in motion in the near future. To make the problem tractable we consider a one-dimensional signal $z(t)$, which is typically the time variation in a single element of the pose trajectory. The problem is, therefore, to develop a mapping $z'(t) = \mathcal{F}[z(t)]$. Motion requires an impulse (force). An impulse will, in general, generate discontinuities in some derivatives of the signal. Locating these discontinuities is straight forward and yields a time τ when the impulse was applied. Not all forces lead to a discontinuity — the case of simple harmonic motion (SHM) being one example. In this case the force changes smoothly over time so that the signal is continuously differentiable; in such cases we use extremes of position to determine a τ.

However a τ is determined we proceed as if an impulse has been applied at that instant. Now we face a subtle problem: as a motion emphasis cue, anticipation is supposed to alter the pose trajectory *before* τ, because the change is to be anticipated; but if we do so the resulting animation appears unconvincing. Our solution is to affect the pose trajectory between τ and $\tau + \delta$. The δ is the duration that anticipation lasts; that is the interval over which we affect the pose trajectory in response to the impulse. We ensure the pose trajectory returns to its true state at the interval end. This gives the visual impression of the impulse being applied a short time after τ so the anticipation cue behaves as expected.
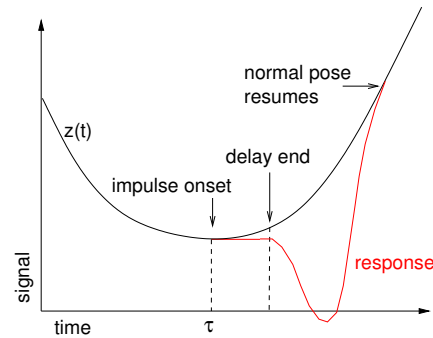

Figure 3: *Illustrating a response to an impulse "patched into" a signal to generate anticipation effects.*

The response has two main parts: a delay and a movement. The delay simply holds the current pose for a short period, the movement is characterised by "deflecting" the pose trajectory in the direction of $-\ddot{z}(\tau)$. This generates a turning point in the response, after which the response "catches up" with the observed pose trajectory; Figure 3 illustrates.

In practice the animator is given control over the magnitude of response as well as temporal duration of the delay, the location of the peak response, and the time it take for the system to return to normal. Figure 4 shows animation frames of a metronome anticipating motion by "snapping" [Wil01]. The bending is due to our deformation effects acting on the modified pose trajectory and indicates inertia, which is why the beater bends as if to oppose motion.

3.2. Motion Exaggeration

Motion exaggeration is another form of dynamic cue. From an animators point of view, motion exaggeration characterises the way an object moves much as a newspaper cartoonist might exaggerate facial features or an impersonator exaggerates vocal idioms. Intuitively, these characteristics are outliers compared to a distribution of common cases — its unusual for a man to have a high-pitched voice, and impersonators may take advantage by exaggeration.

This principle has been put to use to produce cartoon-like versions of a face [LCXS02], as follows. An eigenmodel is generated from mug-shots of many people by considering each images as a vector in some high-dimensional space. An individual mug-shot is projected into this eigenspace, scaled away from the mean, and then reconstructed to reveal a "cartoon". We might proceed by analogy, at least for cyclic motions such as a walk. The set of pose trajectory for a walking motion must lie on a annular manifold embedded within pose space (the space comprising all possible pose vectors). The eigenvectors of this trajectory point in the most important directions. We can scale a pose vector await from the mean, in proportion to the eigenvalues associated with the eigenvectors, thus scaled more along the important

Figure 4: *Frames in an animation showing the instant of a (virtual) impulse; (left), the peak anticipatory response; someway towards returning to normal; and normal motion (right). All motion is subject to a deformation so to make the effect of anticipation clearly visible.*
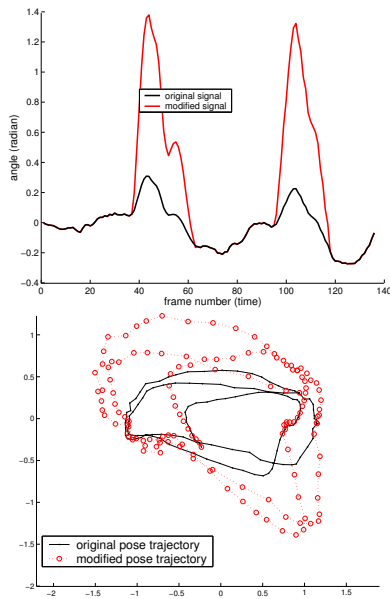

Figure 5: *Left: The change in the pose trajectory of the left upper leg. The original signal (black) is held fixed when then left foot is on the ground, but is altered (red) when the left foot is off the ground. Right: The pose complete trajectory projected into the plane defined by the largest two eigenvectors. The original pose(black) is clearly scaled subject to constraints — some points remain static — to yield an updated pose trajectory (red).*

directions. This approach is poor: (1) The doll can contradict physical constraints, so that feet appear to slide along the floor or look as they *should* penetrate the ground, for example; (2) The output can be aesthetically displeasing — which is difficult to quantify but is important nonetheless; (3) It offers little scope for animator control, which is probably related to point 2. We have only indirect evidence to support this: animators produce output with an aesthetic value greater than any machine can manage at this point in time.

Our approach is to allow animators to impose physical constraints, so that feet are fixed to the ground when necessary, but that the remaining motion is exaggerated by scaling away from some mean. Consider a full pose trajectory $\mathbf{p}(t) \in \Re^n$. Animators are able to specify a subspace that remains can move between times τ_1 and τ_2 using a projection matrix $\mathbf{M}(t) \in \Re^{m \times n}$ that "picks out" those dimensions of the pose trajectory that *can* be changed at some time t. Thus

$$\mathbf{q}(t) = \mathbf{M}(t)\mathbf{p}(t) \tag{12}$$

identifies those elements of pose that can vary at time t. Typically each row of $\mathbf{M}$ is drawn from the n^2 identity matrix. We can now synthesise a new pose vector:

$$\mathbf{p}'(t) = \mathbf{p}(t) + \mathbf{M}^T(t)\mathbf{q}'(t) \tag{13}$$

where $\mathbf{q}'(t) = \mathcal{F}[\mathbf{q}(t)]$ is some modified version of the "variable" pose.

We have found that simply scaling away from the mean of the subspace yields better but nonetheless poor results. This is because scaling along eigenvectors tends to obscure those high-frequency characteristics a walk (say) as individual. Our approach is more subtle. We first transform the signal by $\mathbf{R}$ so that is principle eigenvector aligns with the 'x'-axis: $\mathbf{r}(t) = \mathbf{R}\mathbf{p}(t)$. Next we fit a piecewise curve $\mathbf{s}(t)$ smoothly approximate $\mathbf{r}(t)$. Then we measure the error signal $\mathbf{e}(t) = \mathbf{r}(t) - \mathbf{s}(t)$. We then map as follows:

$$\mathbf{q}'(t) = \mathbf{M}^{-1}(w(t)\mathbf{A}\mathbf{s}(t) + \mathbf{B}\mathbf{e}(t)) \tag{14}$$

where $\mathbf{A}$, $\mathbf{B}$ are linear transforms and $w(t)$ is a smoothing function that ensures the scaling is zero at the edges of the time window $[\tau_1, \tau_2]$. Without this weighting the motion suffers a discontinuity at window boundaries. This approach has the advantage of separating high-frequency detail from low-frequency detail and the effect on a particular signal is shown in Figure 5.

We applied this mechanism to create an animated sequence, stills from which are shown in Figure 6. The pith-helmet and handle-bar moustache were painted using techniques described elsewhere [Col04]. A commercially available product added the 1920's cinematography effects.

Figure 6: *A "Monty-Python" funny-walk; the product of motion exaggeration applied to walking.*

3.3. Futurist Art

Our attempts at synthetic Futurist art is unique in non-photorealistic rendering. The nearest alternative is the automatic production of Cubist art from three or four photographs [CH03]. We began by studying Duchamp's "Nude Descending the Stairs"; painted in response to the work of motion scientist Étienne-Jules Marey [Cab67]. "Nude Descending the Stairs" is a complicated piece of Art, a plethora of arms and legs intertwine and obscure one another; motion blurring, ghosting, streak-lines, and other artifacts usually associated with animation are crammed into the painting. Duchamp succeeds in creating a sense of motion without ever painting a single form that can be recognised as definitively human.

We have discovered that careful analysis of pose trajectory is the key to synthetic Futurist art. More specifically, the motion of the feet can be used to control the whole process. The angle that a foot makes to the lower-leg is, to a first approximation, sinusoidal. The cycles of the feet are in anti-phase. Duchamp used a particular 1/4 cycle of the nearest foot to "cue in" motion blurring, and the corresponding 1/4 cycle of the rear foot to cue streak-lines, shown in Figure 7. The most robust way to identify these partial cycles is to analyse the pose trajectory of the foot and lower-leg — the limbs that are pivoted by an ankle. These 1/4 cycles correspond exactly to those time periods when the relevant foot is not on the floor, as Figure 7 also shows. In fact the start and stop of the cycle corresponds to salient points on the *spatial trajectory* of the ankle. We note that such analysis provides an opportunity to automate motion exaggeration yet further.

Finding the minima and maxima of a pose trajectory is complicated by the fact the signals can be very noisy. Low-pass filtering the signal is not acceptable because it causes temporal movement in the peaks. Influenced by the *sieves* of Harvey and Bangham [BMHF99], we have developed a simple yet robust non-linear filtering of (some element of) a pose trajectory that we have developed specifically for this problem, but which generalises to wider contexts.

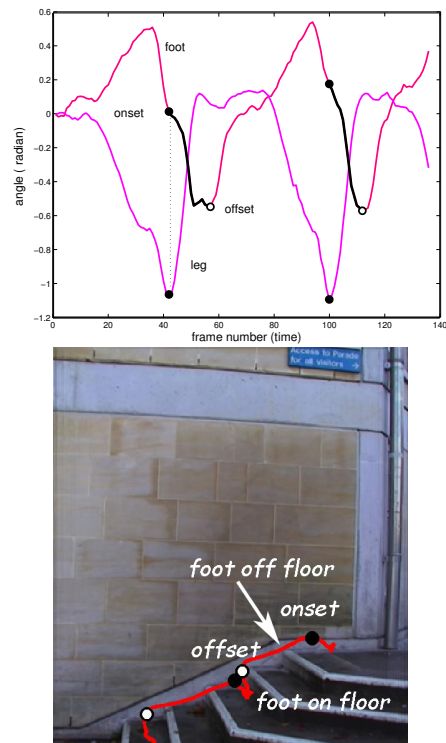Let $z(t)$ be a noisy signal and $d\tau$ be the width of a temporal

Figure 7: *Top: the foot and lower-left pose trajectories are in quadrature, making the 1/4 cycle during which the foot is of off the floor relatively easy to find. Bottom: The spatial trajectory of the near ankle (red) partitioned into sections between the onset (black) and offset (white) of the 1/4 cycle. The foot is fixed to the floor for the remaining period.*

window. The window moves over the signal and the location of turning points are recorded; an extremum at a window boundary is not sufficient to count as a turning point, unless the window boundary is coincident with the trajectory boundary. Appealing to Marr's idea that salient edges persist over scale we posit that dominant turning points persist over scale. Following Witken [Wit83] we construct scale-
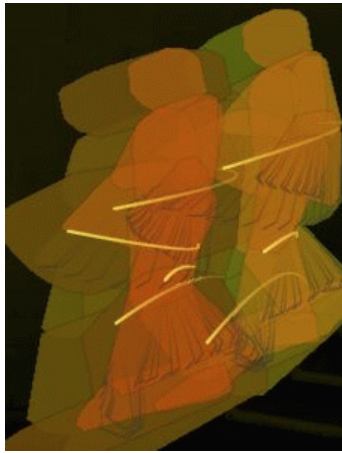
Figure 8: *"Figure descending the steps"*

space trajectories of turning points, choosing persistent maxima the lie between persistent minima. Because there is no modification of the signal there is no shift in the position of extrema.

The motion blur effect Duchamp uses is recreated by "welding" polygons around a limb. As a limb moves through the 1/4 cycle we record the location of its polygon and "weld" these polygons by finding their convex hull. The depth of this amalgamated polygon is fixed at the depth of the contributing limb. When all polygons for all limbs have been amalgamated in this way over the whole time period of the video we acquires a set of depth ordered amalgamated polygons. These are rendered in back-to-front order. By making the polygons partly transparent the visual effect is to entwine the limbs, yet the near polygons appear brighter so that so visual sense can be discerned from the picture.

Ghosting and streak-lines are produced using techniques described elsewhere [CRH03]. Ghosting marks are painted on the near limbs only — rendered on top of the welded polygons. Streak-lines are traces of the rear polygons, clipped against the welded polygons of the rear limbs but painted as the top-most layer. The resulting Futurist artwork is shown in Figure 8.

4. Concluding remarks

This paper described our initial steps towards automatically synthesising dynamic cues from video, focusing on anticipation and motion exaggeration. Whether the principles we have introduced in addressing these case generalise easily is unknown. It is likely that inverse kinematics of some kind will play a major role in automating anticipation, although whether pose analysis will ever be of sufficient power to produce the necessary key-frames is an open problem.

As presented, our framework for dynamic cues is premised on the automatic recovery of articulated struc-

tures. However initial experiments operated at a lower level of abstraction, requiring no such model and allowing limbs to move without the constraints of pivots. Although more generally applicable, the aesthetics of the resulting motion were disappointing. It is likely that the conceptually higher level model of the articulated structure confers more believable movement because it more closely matches our mental model of the way in which our subjects move. By substituting our hierarchical model with, say, a facial muscle model, we may be able to create anticipation in alternative classes of subject commonly used by animators. Future work might address a methodology for the selection and substitution of models by the computer animator.

Open questions notwithstanding, we have introduced a number of useful analysis techniques: automatic inference of articulated structure under planar motion; constrained scaling of pose in eigenspace; a robust signal filter to locate turning points. The dynamic cues synthesised by this framework have been integrated into a larger "Video Paintbox" system (see [Col04]), capable of both alternative motion emphasis styles (through augmentation and deformation cues) and also flicker-free visual stylisation of content (for example, cartoon shading and painting). In addition, our initial experiments in the emulation of Futurist artwork point toward interesting possibilities for study in NPR, with respect to generating both static depictions of motion and abstract artistic styles. We had not anticipated that a simple analytic explanation might lie behind Duchamp's artwork, and this has certainly added to our appreciation of it.

References

[AHSS04] AGARWALA A., HERTZMANN A., SALESIN D., SEITZ S.: Keyframe-based tracking for rotoscoping and animation. In *Proc. ACM SIGGRAPH* (2004), pp. 584–591.

[BH00] BRAND M., HERTZMANN A.: Style machines. In *ACM SIGGRAPH* (2000), pp. 183–192.

[BLCD02] BREGLER C., LOEB L., CHUANG E., DESHPANDE H.: Turning to the masters: motion capturing cartoons. In *Proc. 29th Intl. Conference on Computer Graphics and Interactive Techniques (ACM SIGGRAPH)* (July, 2002), pp. 399–407.

[BMHF99] BANGHAM A., MORAVEC K., HARVEY R., FISHER M.: Scale-space trees and applications as filters... In 10th *British Machine Vision Conference* (1999).

[Cab67] CABANNE P.: Entretien avec marcel duchamp, 1967.

[CH03] COLLOMOSSE J. P., HALL P. M.: Cubist style rendering from photographs. *IEEE Transactions on Visualization and Computer Graphics 9*, 4 (Oct. 2003), 443–453.

[Col04] COLLOMOSSE J. P.: *Higher Level Techniques for*

the Artistic Rendering of Images and Video. PhD thesis, University of Bath, U.K., May 2004.

[CPIS02] CHENNEY S., PINGEL M., IVERSON R., SZY-MANSKI M.: Simulating cartoon style animation. In *Proc. 2nd ACM Symposium on Non-photorealistic Animation and Rendering* (June 2002), pp. 133–138.

[CRH03] COLLOMOSSE J. P., ROWNTREE D., HALL P. M.: Video analysis for cartoon-like special effects. In *Proc. British Machine Vision Conference (BMVC)* (Norwich, U.K., Sept. 2003), vol. 2, pp. 749–758.

[CRH05] COLLOMOSSE J. P., ROWNTREE D., HALL P. M.: Stroke surfaces: Temporally coherent artistic animations from video. *IEEE Trans. Visualization and Computer Graphics 11*, 4 (2005).

[Hic03] HICKS Y.: *Modelling and tracking of articulated human motion.* PhD thesis, University of Wales, Cardiff, 2003.

[HP00] HERTZMANN A., PERLIN K.: Painterly rendering for video and interaction. In *Proc. 1st ACM Symposium on Non-photorealistic Animation and Rendering* (2000), pp. 7–12.

[Kal60] KALMAN R. E.: A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering 82* (Mar. 1960), 35–45.

[LCXS02] LIANG L., CHEN H., XU Y.-Q., SHUM H.-Y.: Example-based caricature generation with exaggeration. In *Proc. Pacific Conf. on Graphics and Applications* (2002), pp. 386–393.

[Lit97] LITWINOWICZ P.: Processing images and video for an impressionist effect. In *Proc. 24th Intl. Conference on Computer Graphics and Interactive Techniques (ACM SIGGRAPH)* (Los Angeles, USA, 1997), pp. 407–414.

[SPR*94] STROTHOTTE T., PREIM B., RAAB A., SCHUMANN J., FORSEY D. R.: How to render frames and influence people. In *Proc. Computer Graphics Forum (Eurographics)* (Oslo, Norway, 1994), vol. 13, pp. C455–C466.

[Wil01] WILLIAMS R.: *The Animator's Survival Kit.* Faber, 2001. ISBN: 0-571-21268-9.

[Wit83] WITKEN A.: Scale-space filtering. In *Proc. Int. Joint Conf. on Artificial Intelligence* (Karlsruhe, Germany, 1983).

[WXSC04] WANG J., XU Y., SHUM H.-Y., COHEN M.: Video tooning. In *Proc. ACM SIGGRAPH* (2004), pp. 574–583.