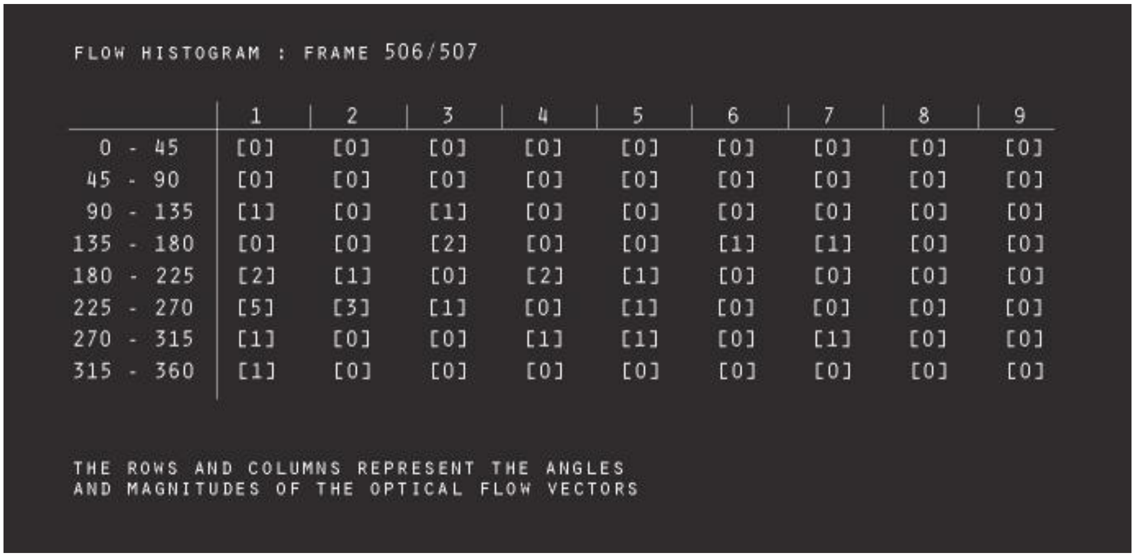
GESTUS

Hector Rodriguez

Gestus is a moving image analysis and processing framework that explores the relationship between algorithmic procedure and symbolic form. The core technical and aesthetic concept is the vector, understood as a method of representation or symbolic form that expresses an abstraction of movement. Its aesthetic effects are best described via the vocabulary of cognitivist aesthetics, as the posing of a perceptual challenge to an active viewer.



```
FLOW HISTOGRAM : FRAME 506/507

            |  1  |  2  |  3  |  4  |  5  |  6  |  7  |  8  |  9
------------------------------------------------------------------
  0 -  45   | [0]   [0]   [0]   [0]   [0]   [0]   [0]   [0]   [0]
 45 -  90   | [0]   [0]   [0]   [0]   [0]   [0]   [0]   [0]   [0]
 90 - 135   | [1]   [0]   [1]   [0]   [0]   [0]   [0]   [0]   [0]
135 - 180   | [0]   [0]   [2]   [0]   [0]   [1]   [1]   [0]   [0]
180 - 225   | [2]   [1]   [0]   [2]   [1]   [0]   [0]   [0]   [0]
225 - 270   | [5]   [3]   [1]   [0]   [1]   [0]   [0]   [0]   [0]
270 - 315   | [1]   [0]   [0]   [1]   [1]   [0]   [1]   [0]   [0]
315 - 360   | [1]   [0]   [0]   [0]   [0]   [0]   [0]   [0]   [0]
            |

THE ROWS AND COLUMNS REPRESENT THE ANGLES
AND MAGNITUDES OF THE OPTICAL FLOW VECTORS
```

Fig. 1. Flow Histogram representing the magnitudes and directions of a set of flow vectors.

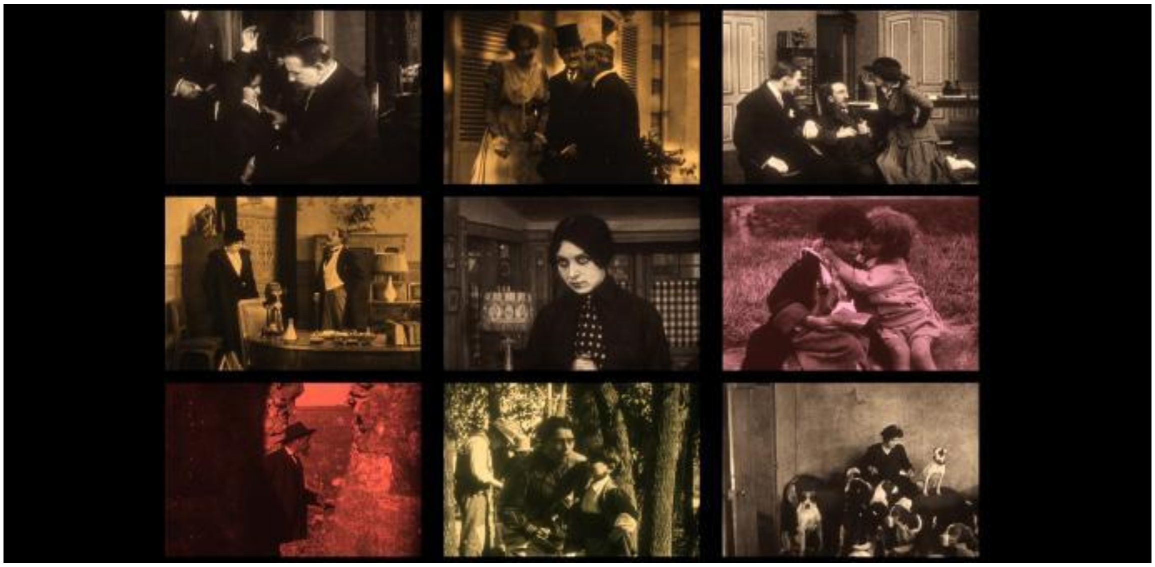


Fig. 2. Matrix display showing a video clip (center) surrounded by the eight best matches.

Vectoral Form

Experimental filmmaking can be seen as a radical critique of the conventions of linearity and transparency that characterize the conventions of “classical” narrative cinema. Classical conventions organize stylistic parameters around the clear and consistent communication of story information. The image is framed to direct the viewer's gaze to the main points of interest, relative to the main line of action, which supposedly unfolds in a coherent spatiotemporal domain and advances mainly through confrontations between goal-oriented agents. This dominant system always binds movement to objects and places. In this respect, classical narration draws on well-established features of ordinary cognition. Our awareness of movement is typically bound to specific objects and locations. We normally see (not movement as such but rather) some*thing* moving some*where*.

Experimental film and video makes have explored alternative modes of narrative organization and spectatorial address. In particular, there is a strand of avant-garde cinema that draws on the power of formal abstraction. Filmmaker Hollis Frampton, for instance, has called for “progressively more complex a priori schemes to generate the various parameters of film-making...” [1] The aesthetic potential of computational media lies precisely in its power to generate abstractions and so to extend this artistic lineage. This essay describes a particular kind of abstraction, which I call “vectoral form”, and its potential utility as a method of avant-garde production.

The vector concept is here understood as a symbolic form, a method of representation. A pervasive feature of the cybernetic society, the vector plays a fundamental role in such control and surveillance tasks as motion tracking, action recognition, abnormal behavior detection, and video compression. These tasks substantially depend on algorithms that estimate the movements occurring in some image stream and then represent them as vector fields. A vector is an abstraction of movement. It is essentially characterized by two properties, magnitude and direction, and is often visualized as an arrow of a certain length and orientation. Vectoral form provides a sort of common currency that renders distinct movements quantitatively commensurable. It affords the possibility of measuring the similarity between motions by comparing the magnitudes and directions of their respective vectors. I claim that this abstraction potentially supplies the media artist with a radical principle of formal organization.

This essay explores the artistic possibilities of vectoral form in the context of the *Gestus*framework, a custom software system designed for the analysis and re-assemblage of video data. The system uses vector representations to search for similar movements occurring in one or more movies, and then displays those motions side by side as a split screen or multiple-channel projection. This emphasis on movement-as-such grows out of the technical role played by vector representation in the production process. The formal-aesthetic characteristics of the work are thus derived from the algorithmic principles that produced it. To quote filmmaker Malcolm LeGrice, *Gestus* tackles “the question of procedure as a determinant of form.” [2] The technical procedure used here involves several steps.

Abstracting Movement

Assume that one or more “black and white” (grayscale) films are already available in digital form. Each film is represented as a sequence of frames and each frame consists of a two-dimensional array of pixels. The color of each pixel is represented as a floating point number in the interval (0,255). The first step in the algorithm normalizes the data by subtracting from the value of each pixel the mean (computed over all the pixels in one frame) and then dividing by the standard deviation.

The next step deploys an optical flow technique, in this case the Lucas-Kanade (LK) algorithm, to estimate the motion between each pair of successive frames. [3] LK assumes that clusters of contiguous pixels move together as a whole from one image to the next. Take for instance a close-up of a person's face. The eye portion will occupy several pixels, which tend to move together as a single group from frame to frame. The algorithm implements this assumption by partitioning each image frame into rectangular windows or "flowpoints", each of which is then tracked as a coherent group. The output of the algorithm is a field of motion vectors that estimate the displacement of each flowpoint from one frame to the next.

The optical flow for the ith frame gives an estimate of the flow from the nth to the (n + 1)th frame. The set of optical flow vectors for a single frame can be represented as

$$Fi = \{fi1, fi2, ..., fij , ..., fiJ\}$$

where

$$fij = [Xij, Yij, \vartheta ij, Sij],$$

such that the coordinate (Xij ,Yij) represents the on-screen location of each vector, and ϑij and Sij represent the orientation and velocity of the flow vector for that location. Optical flow data therefore consists of *bound* vectors, each associated with a specific position on the frame. A vector is bound if it has a definite location, which can be described numerically via (e.g.) screen coordinates.

The algorithm then unbinds these vectors by abstracting away all location information. Flow data is quantized into $Nv \times N\vartheta$ bins, forming a two-dimensional matrix or *flow histogram* (FH), which will be subsequently used as the basis for comparison. The (i,j)th entry of the matrix represents the number of vectors with magnitude i and direction j. (Fig. 1).The algorithm will henceforth proceed solely on the basis of the magnitudes and orientations of the vectors, not their coordinate positions on the screen, effectively treating all vectors as free (unbound) vectors.

Each FH is then further processed (in technical terms, the algorithm performs a Principal Component Analysis, and selects the top N eigenvectors) and transformed into a descriptor $\boldsymbol{x} = (x1, x2, ..., xn)$. I shall refer to these descriptors as "motion frame projections" (MFPs). Each MFP gives a highly abstract representation of the movement between two frames. The dissimilarity ("distance") between any two MFPs x and y can now be defined as follows:

$$\mathrm{dist}(x,y) = (x1 - y1)2 + (x2 - y2)2 + \ldots + (xn. - yn) 2$$

This measure expresses a quantitative comparison between the movements of any two pairs of frames. It is straightforward to compute the distance between two video segments, each consisting of an arbitrary number of frames, simply by computing the distances between each corresponding MFP and then adding them together. Two segments "match" if the distance between them is sufficiently small.

After obtaining MFPs for each frame in the film(s) to be processed, the algorithm groups them into short "matching segments" of fixed length. My first experiments used L = 2 frames (1 MFP), but it is more perceptually rewarding to compare movements that extend over several frames, so I settled on a fixed length L = 10 frames (9 MFPs). This length determines what counts as a single "instant" or "gesture", from the standpoint of the searching and matching algorithm. It is then possible to select any arbitrary segment S0 of length L and search for other segments S0, S1, S2, ….of length L (which can but need not belong to the same movie) that closely match it. In the current version of this project, we select the 8 “best” (closest) matches, displayed as a matrix around S0, which occupies the central cell of a 3 x 3 grid (Fig. 2).

Aesthetic Effects

The chosen source material is Louis Feuillade’s 1916 film serial Judex. There are several reasons that justify this choice. Feuillade worked within a tradition of ‘tableau cinema’ that relied on deep space staging rather than camera movement or analytical editing. Film theorist and historian David Bordwell stresses the director’s skill at forming dynamically changing geometric arrangements of bodies in space, carefully directing the viewer’s gaze to salient features of a scene on a moment-to-moment basis. “Such gentle geometries of movement hard to find in today’s cinema, and observing them in Feuillade reminds us that long ago some directors crafted their images as two-dimensional patterns of bodies in space.” [4] Bordwell has noted the rhythmic quality of cinematic motion in Feuillade’s work: “Shots are subtly balanced, then unbalanced, then rebalanced...” [5] By focusing attention purely on the magnitude and direction of movement, *Gestus*foregrounds the rhythmic quality of Feuillade’s deep space orchestrations.

The multi-channel display cues the viewer to engage in an active process of visual thinking, scanning the various images in an effort to identify the similarities between them. Her perceptual effort becomes an integral element of the vector machine. The system might display a human hand alongside a bird’s face, for instance, thus revealing the kinetic resemblances of otherwise heterogeneous objects. This interplay of similarity and difference underpins the main aesthetic effects of the *Gestus* system, its visual demonstration of the difference between movement and the thing that moves.

Sometimes, the viewer easily detects similarities between the various images. In other cases, however, the movements are very subtle and occur in different areas of a crowded image, posing a sharper perceptual challenge. Perhaps a dropping hand near the bottom of one image corresponds to a leaning shoulder near the left edge of another. The system invites, challenges, and sometimes frustrates the spectator’s cognitive-perceptual skills. The gaze is made restless. Although the software uses segments of fixed length, the moment-by-moment experience of lived duration sometimes expands or contracts, depending on the effort required to bring the various images into perceptual relation. This destabilization of the gaze demonstrates the transgressive possibilities unleashed by the abstraction of movement through vectoral form. Motion tracking techniques designed for surveillance and control are thus detourned and redeployed.

Liberate the vector.

References and Notes:

1. *Quoted in James Peterson, Dreams of Chaos, Visions of Order: Understanding the American Avant-Garde Cinema (Detroit, MI: Wayne State University Press, 1994), 112. This book also provides a good introduction to cognitive aesthetics.*
2. *Malcolm LeGrice, Abstract Film and Beyond (Cambridge, MA: The MIT Press, 1977), 88.*
3. *Bruce D. Lucas and Takeo Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," Proceedings of Imaging Understanding Workshop, Washington, DC, (April 1981), http://cseweb.ucsd.edu/classes/sp02/cse252/lucaskanade81.pdf (accessed June 8, 2012).*
4. *David Bordwell, "Revising Our Sense of Feuillade," David Bordwell's Web Site on Cinema, 2005, http://www.davidbordwell.net/books/figures_intro.php?ss=2 (accessed September 1, 2011).*
5. *David Bordwell, "How to Watch Fantomas and Why," David Bordwell's Web Site on Cinema, 2005, http://www.davidbordwell.net/blog/2010/11/11/how-to-watch-fantomas-and-why/ (accessed September 1, 2011).*