Written as an artist's talk for Interaccess (a Toronto artist-run media gallery), 2008 SEQ CHAPTER \h \r 1

Digital Cinema

R. Bruce Elder

For the past several years I have been working in digital cinema – that is, cinema that incorporates digital images. To be able to do this, I went back to night-school and, I suppose, did the equivalent of a degree in applied mathematics and computer science – and devoted much of research efforts to writing computer programs for processing images in ways that conform to my aesthetic ideas. There is, I believe, a radical breach between the classic, photographically-based cinema and new, digital cinema. Of a photograph, it is always reasonable to ask: "Who created the photograph, the photographer or nature?" The question cannot be answered, but one must acknowledge that an aspect of the beauty of a photograph, is that a photograph, as André Bazin pointed out long ago, strikes us as a phenomenon of nature. Thus, making photographs, or photographically-based films, seemed to me a way of cherishing the gift of what reality – actually, of what is more accurately described as *natura naturans* – makes for us. Photography, I concluded, succeeds when the photographer transcends wilfulness and learns to cherish the gifts given him or her. That the will of the human "maker" should have no place in the creation of the image is the most radical implication of photography, its most profound rupture with the traditional image-making. Photography even allows the imagination to be circumvented, and by this, it reveals our being-with-the-world

Digital cinema is another matter entirely – digital cinema gives the subject back its traditional role – or, rather, something close to its traditional role. The imagination resumes it traditional function in image-making: digital images once again require that we step back from the world and enter into the space of subjectivity. And all the usual ontological and epistemological problems that tradition has with images – both philosophical tradition that descends from Plato and the theological tradition that developed out the Judaic religion – reassert themselves with digital images: one can ask whether what one sees in the image is an object; the image allows for contradictory interpretations, and so exposes the knowledge we gain through them to doubt; images mediate between the subjective and the objective world (and so, in a process that Baudrillard's writings expose, take precedence over the objects they purport to represent and, finally, block access to those objects).

The digital image does offer something that changes the tradition of image-making: it makes it possible to realize the Pythagorean dream of producing images/reality through number and through calculations of a complexity that Pythagoreans could never have conceived them rather than through the depictions of facts (pictures as Wittgenstein understood them – arrangements of elements that mirror states of affairs). This is a whole new possibility for the imagination, and, though I have dabbled with it, I have not been able to come to terms with it.

Insisting on the role of subjectivity and imagination in the production of digital images of course raises the question of the subject, and issues around the subject are vexed. One can see the appeal of the idea that the subject is something that can be detached from one body and transplanted. For let's admit that the term "I," as it is ordinarily used in intellectual discourse, is hopelessly troubled – its meaning all but indiscernible, inasmuch as it is surrounded by a thick fog of philosophical, theological and psychological confusions. The new technologies, and the new media they have produced, promise to help render this traditional confusions obsolete, for they propose a new definition of the subject: “I” am a complex system of electromagnetic and chemical brain processes. This new start on the description of the subject at least promises to clarify what this reality that we refer as “I” really is.

But it also proposes the hope that the “I” might be transplanted – already the brains (or parts of brains) of rats can transplanted from dying bodies to fetuses. In this sense, the immortality of “I,” through the repeated transfer of brain parts, has become a theoretical possibility that the new technology will undoubtedly strive to realize.

Despite all the confusions that have surrounded the traditional concept of the subject, and potential of the new conception to sweep away that fog of confusions, I find the whole idea really pernicious. It denies the important role of the particular body that each of has in establishing his or her identity – thus, because our identities are so crucial, it devalues the body. All my recent work – and much of the work that I did earlier (though I didn’t realize it at the time) – has been in devoted to enhancing the sensation of flesh (and flesh’s belonging to the world). I think the topic of the body is the most important topic one can devote oneself to in this contemporary climate, where “despisers of the body” are so prevalent. If I were starting out now (or, rather, if were young enough) I would surely be doing very “in your face” performance pieces that dealt with the body. They would be “in your face” pieces not so as to be transgressive – the idea of transgressive art strikes me as among cultural theory's most boring ideas – but because body art can so bring to our attention the importance of rapture. By “rapture” I mean any intense experience (because of its extreme intensity, such experience is sometimes is felt as unpleasure) that deranges focussed, analytical consciousness – experience so intense that it leaves conventional ways of thinking in ruin (at whatever cost). This sort of experience invariably makes us sense the body’s role of experience: we feel at once acute anxiety (one can even feel a some measure of nausea) and acute pleasure as every nerve ending seems to tingle and we feel waves of bliss surge through the body, from head to toe and from toe to head. One experiences this when the energy of the body rises up and imposes itself on us, as occurs in love-making, or when we find ourselves intensely aroused and intensely embarrassed at the same time – I sometimes experienced it when my assistant photographed me for films we made together (as she did many times), and it was primarily those occasions that taught me the crucial importance of this sort of experience. I am sure that making performance art would allow me to focus much more directly and clearly on that sort of experience, which I believe is so important.

The most important thing that the body teaches us through experience of this sort is that we are "owned" – first by the divine, that fills flesh with desire, and then, through the divine, by all other people. Thus flesh teaches us we owe deep allegiance to one another. It teaches us that the obligation that any other person imposes on me, just by being human, is absolute and unconditional; and we have even more profound obligations to those who fall into the circle of our love. I have no choice but to care for others around me, and seek to life a life in which I care profoundly for a number of friends whom I love deeply. Through the divine, they own me; and I am not free to choose what I wish to do – I belong to them and must act out of my concern for them.

Art should reveal areas of experience that we cannot reveal to others except by making art – forms of imagining that we cannot reveal in any other way. By doing this, art gives us a more profound sense of what human be-ing is – it acknowledges that we think in ways other than the analytic/instrumental/propositional forms that have dominated us (in the West) at least from the time of the Enlightenment (and possibly earlier). Propositional thinking, thinking that can be modelled by rewriting of one string of symbols into another according to an established rule (thinking of the sort that computer scientists are prone to think of as the sole form of thinking available to us), is "deranged" by the intense feelings surge through the body as it responds to the field of energy that lies "beyond us." I believe it is important to tell one another that we possess the capacity for rapture, too – and for all those sorts of experience that are close to rapture (such as mad love or the states that strong, repetitive rhythms induce, a state akin to trance and prayer). To do this, art has to acknowledge the less seemly contents of our thoughts.

The fundamental responsibility artists have is to make contact that beneficent field of energy that lies around us and to enter into that energy. When one does this, it takes control. Then one's responsibility is to obey its commands; in this, there is no "freedom of imagination." One takes orders—a n image flashes into your mind and you have to make it, no matter how wrong-headed or embarrassing or unaesthetic or humiliating it might be. One cannot be allow considerations of audience/reception to intrude upon this; such concerns make one less willing to go to the extremes to which one might be commanded to go – one might get to thinking "What will people think?!?." When one becomes disobedient, one breaks with the source of one's imaginings. This is why the Muses were frequently thought to be very jealous.

I am concerned to reject assertions like those of Arthur Kroker that the new media (cyberspace and virtual reality) will lead us into a better future – a future that will undo all the devastations of the centuries since the Enlightenment have wreaked on us, a future that will be heaven-on-earth. One hysterical comment from Kroker, formulated on the model of neo-Platonic philosophers Robert Grosseteste's metaphysical light proposes "So begins our violent descent into the electronic cage of virtual reality. Down we go into the floating world of liquid media where the body is daily downloaded into the floating world of the net, where data is the real, and where high technology can fulfill its destiny of an out-of-body experience." Gibson opines that soon we be shuffled off into "bodiless exultation." And Microsoft asks "*Where do you want to go today?*" as we sit in front of the screen of our monitor. Eric Voegelin, in his *The New Science of Politics* identifies the historical shift that generated these hysterical, and body-despising, comments: in the 13th Century, Joachim of Flora (or Fiore) broke with the Augustinian notion of a de-divinized "Civitas Dei" by resurrecting the Gnostic notion of heaven-on-earth. Joachim was was nominally a Cisterian monk in Calabria but actually a Gnostic. One of Joachim's contribution to the history of millenarianism was the notion that history should be divided into three periods that correspond to the three persons of the Trinity. The Second Age of the Son was coming to a close, Joachim professed, and the glorious Third Age of the Spirit was about to dawn.

Later Utopian movements adopted this formula for dividing history into three periods. Ivan IV forced Constantinople to recognize Moscow as the Third Rome in 1589 – an early painting of the Theosophist/Gnostic painter Wassily Kandinsky in fact depicted the dream that Moscow would be the Third Rome. The historical fantasy that was the Third Reich incorporated possessed the same mythological structure. The later example, especially, imposes on new media thinkers who adopt the gnostic metaphysics – and they are many – the responsibility of considering with whom else (besides the Urantians, Tim Learyian reprogrammers, and other extravagant cult-adherents who have played a role in formulating the received "metaphysics of digital reality") they are associating themselves

According to one prevalent conception of the metaphysics of digital reality, the convergence of the media (of text, image, moving image, and sound, all "interactively" available) promises to unite non-corporeal information and non-corporeal individuals in the same electronic medium, in which everything and everybody are co-extensive. This total co-extensivity is the basis for the “total awareness” my new media students keep telling me is dawning (or rather, I understand, their classes inform them is dawning). This idea of the non-corporeal self, of the self that is identical with information, is a modern version of the soteriological dream of transcendence through the emptying out of the self. The appeal that the idea of dematerialization has to new media theorists is that it supposedly exposes that nothing possesses an internal principle that accounts for its growth – that the self, to take it as an instance, is wholly and completely malleable, and can – and is – constantly made and remade by changes in the conditions of the system of representation that shape it. The Gnosticism of this conception is evident: our world is a wrong world not only because it is a bad world, but also because it offers the illusion of corporeality (that things have an nature by virtue of their constitution). According to the soteriological principles of these new media theorists, why it is so important to see through the illusion of the self – why it is so important to understand that we possess no internal principle but are subject to endless remaking – is that the new non-corporeal world can come under our complete control, because we know how we made it and how to reproduce it. In the end, we would act as a new Creator – this is the dream that fuels those who proclaim that the new media offer unlimited creative freedom, that we might usurp the place of the Divine. We are unshackled from all moral limitations of our world as it is, and nothing outside of us limits our capacity to impose on the world.

The great Canadian philosopher George Grant critiqued this very position in such stunning books as *Technology and Empire* and *Technology and Justice*. Grant showed that the belief that the Good is not inherent in the order of nature underpins that belief, essential to the regime of technique in which we exist and through which we conceive the world, that humans are free to remake the world. Grant pointed out the notion of technique is central to modern civilization – so much so that the progress of techniques has now become the horizon for those who seek to understand the Good. Moderns have lost the ability to understand the standards of goodness by which particular techniques may be judged. The conviction that human knowledge has the purpose of mastering human and non-human nature is central to moderns' ideas about the nature of human being. The idea that new media theorists expound, that human being possesses no inherent nature has the purpose of justifying the proposition that humans can be made and remade at will – that nothing in the nature of human being limits society's/ideology's/the artist's freedom to refashion them. And that conception, in its turn, belongs to a discourse on value and freedom that is associated with the will to technique – indeed it is part and parcel of the modern belief that nature, since it as objectively devoid of value, can be remade at will.

What more than anything impresses me about what the propositions issued as the metaphysics of digital reality is their tendency towards imperial aggrandizement. The consequence of this, I fear, may well be tyranny. I mean "tyranny" here in the Straussian sense, as it arose within a remarkable exchange between Leo Strauss, the renown conservative political philosopher, and Alexandre Kojève, France's great interpreter of Hegel. A key topic of the debate was Kojève's affirmation that "that the universal and homogeneous state is the best social order, and that mankind advances to the establishment of such an order." Kojève pointed out that the final stage of civilization, the establishment of the universal and homogeneous state, comes into being as the secularization of the political ideal of the Christian community, which proposed that all humans could transcend their given differences through their faith, and be made one in the body of Christ's church – I hope everyone noted that this claim resonates in the beliefs of the new media communitarians. Behind this lies the assumption (not unlike that of soteriological assumptions that undergird the metaphysics of digital media), that thought (and specifically, for the ancients, philosophy) takes its bearings not from an ahistorical eternal order, but from eternity as the totality of all historical epochs (the sum of all knowledge that our new hypertextual “koran” represents).

Strauss argued, against Kojève, that the goal of Hegel’s state, universal happiness, is unachievable – and what is worse, that it will end in tyranny. I don’t find myself in agreement with much in Strauss’s political outlook, but on this matter I think he absolutely right – his thesis turned out, in fact, to be prophetic. Strauss’ argument was founded in the classical belief that humans find their fulfilment in that thinking which leads to wisdom – a premise the Hegel had rejected for the premise that humans find adequate fulfilment in that form of recognition that is available to all. Hegel’s gambit, Strauss argued, had effectively lowered the goal of political action, for his idea of universal recognition as the basis of community and state cannot recognize the inevitable differences among humans, and conceives of communities as nexûs of undifferentiated humans. When we must all be the same, no person will be a true thinker. Philosophy will disappear in such state, through the wedding of technology and ideology (a process that probably is now too far advanced to be reversed). The ideas of a totalization of truth and of total awareness (acquired through the complete co-extensivity of the decorporealized mind and the decorporealized text) that cyberspace promises will surely eventuate in tyranny.

It is time to put away this myth of decorporealization, of the totalization of knowledge that will bring history to end. Because it is grounded in the myth of total identity, total transparency, the prevalent metaphysics of digital reality neglects the actual condition of knowledge: it arises from the Gnostic belief in the possibility of immanentizing of the *eschaton*, a belief that goes hand in hand with the idea that the future can be foreseen and planned. The prevalent metaphysics of digital reality is simply the “dream world” of Gnostic lore, where the structure of reality is disregarded, the facts ignored, and the openness of history replaced by a revolutionary step into the New Age. To replace this myth, may I suggest that we return to where all true understanding starts – with the real body, not the amalgam of metal and flesh that is the cyborg nor the data body of Kroker’s Gnostic dream, but the real body of flesh.

Attunement to the rhythm of what unfolds beyond us – a rhythm that is flexible and ever changing, has the strength to release us from the tyranny of an abstract, rationalized temporality. Awareness of rhythm, because rhythm is experienced corporeally, also undoes the effects of the rationalization of space into a wholly abstract form. Contemporary virtual existence has rendered space wholly abstract. The etiology of that form of space can be readily charted, beginning with the geometric optics of the Renaissance. The development of geometric optics and camera obscura led to the rationalization of vision around an axis consisting of the fiction of a single, fixed vantage point outside the depicted scene, at a place established by the vertex of a pyramid, whose base is the surface of the painting and the slope of whose sides is arbitrary. Thus, the body was removed from the scene of vision. But in the nineteenth century representation took on a different character: the space of a drawing, especially those drawings whose primary purpose is to provide information about reality, came to be understood as a Cartesian plane, and the relations between elements in the drawing were to be determined not through appearance, as projective geometry had attempted to do, but rather through measurements, which were then transposed orthogonally to the drawing surface. If the body had been excluded in the system of Renaissance perspective, the subject was excluded in the representational regime that developed in the nineteenth century. When the subject is given no place, the drawing surface itself becomes utopian. That utopic space is the predecessor of the utopia of cyber-nonreality – a non-place where “there is no there there,” and, above all, no place for the body. Paul Virilio points out that cyberspace constitutes a new space without the usual space-time coordinates; as a result, cyberspace engenerders a disorienting and disembodying form of experience in which communication and interaction takes place instantaneously in a new global time, overcoming boundaries of time and space. It is a disembodied space without fixed coordinates, a space in which one loses connection with one’s body, with nature, and with one’s community. It is a dematerialized and abstract realm in which cybernauts can become lost in space and divorced from their bodies and social world. To counter the abstraction of space and time, we insist on working methods that, in their intensity, leave the trace of the body all over them.

SEQ CHAPTER \h \r 1**Algorithms and Chance: an Interview with R. Bruce Elder**
interview by Vicky Chainey Gagnon
(originally published in a newsletter of the Liaison of Independent Filmmakers of Toronto, 2003)

Your most recent film, *Eros & Wonder* was created using digital technology. How did you become interested in using digital technology in your filmmaking.

B: My interest in computers started early, and grew out of my fascination with the fact that beautiful patterns are often mathematically elegant. There is an entire field of design that explores the beauty of mathematical patterns, and I was fascinated by it; from the time I was boy, I read about the Fibonacci series, and the golden mean, and logarithmic spirals – various topics of that sort. There is an another sizable field of investigation, this one rather flaky (to be sure),
known as spiritual geometry, which uses the mathematics of harmony and both an image of and a means for tuning the soul. I spent a lot of reading in that disreputable field of well. You might be surprised how many artists of the last 100 years have.

Your use of digital technology in *Eros and Wonder* is quite different, however. There you used computer technology to process digital images. I believe that you write the computer programs you use to make your films. Could you tell us about these programs?

B: When I decided to use digital processes in my art making, I started by studying the requisite fields of mathematics and computer science; I went back to night-school and took classes for engineers. Working with the knowledge I was able to garner, I

developed a computer application that would allow me to collaborate with the machine to produce “visual compositions”
-- that would allow me to use many of the same principles that I have employed in my filmmaking to day, but would help eliminate subjective whim.

I developed a rudimentary application that stored a set of images into a database along with a set of image descriptors (“meta-data”) and a set of image processing algorithms. The application’s function was to decide what image-processing methods to apply to he images in the database, and to apply them. At first the method for selecting the processing methods to be applied to the
images was pretty simple: Images were partitioned in groups based on the similarities indicated by their descriptors, as were the image processing methods (my decision on which methods most closely resembled other methods was completely informal and subjective); the image processing methods to be applied to a reference image were chosen at random – the operator got to approve the selection, and if he or she approved it, then the methods most similar to the randomly chosen method were applied to the images in the database that most resembled the reference image.

I used this application in a film a finished almost three years ago now, *Crack, Brutal Grief*,
This way of using image processing methods in film/video production interested me enough (and, I thought, the results were good enough) that I wanted to work further on this application.

It was obvious what refinement I should introduce first: using image descriptors as I did was awkward and introduced an unnecessary subjective element that conflicted with the ideal of avoiding authorial imposition. I quickly realised the application would need to use
of methods to “compute” the similarity between the two images algorithmically.

Can you elaborate on the ideal of avoiding authorial imposition?

B: John Cage protested against the idea that an artwork is the product of an artist’s feelings, believing instead that the creative process should imitate nature in its manner of operation. Cage was among the first composer to make the use of chance operations central to his compositional processes -- and he developed a variety of aleatory techniques that allow chance and indeterminacy to play key roles in shaping musical result. Cage insisted that aleatory operations mimicked natural processes and that by imitating the operation of natural processes, the composer bypass his or her limiting ego and allow a larger system or set of systems to shape the work. This principle has been very important to me. The richness of Cage’s writing helped make the use of aleatory techniques common among composers. The rigour of writings by Iannis Xenakis and James Tenney -- composers who, like Cage, took an interest in stochastic methods – and the power of their works re-enforced this influence.

Over the past few years, I have worked on projects that explored the possibility of extending these composers’ ideas to the visual domain. The initial framework for this exploration was drawn from composer James Tenney who made extensive use of measures of similarity in the analysis of music structures in his book *Meta+Hodos*. I was

intrigued by the possibility of
developing analogous compositional procedures for working with sets of images and, in particular, by the possibility of using measures of similarity to constrain random processes.

How does your computer programme calculate image similarity?

B: The process takes place in a number of steps:
1) Load the “key image” or “query image” (the image for which we want to find similar images).
2) Utilizing methods of feature extraction, measure a number of features of the key image. This stage creates a “signature” for the image.
3) For every image in the database, load and generate a signature. 4) Calculate the Euclidean distance between the signature for the key image and the signatures for each of the database images. Sort and store these values -- what results is a list that shows the proximity (based on its signature) of each of the database images to the query image.

The features I used for creating an image signatures were the intensity of the image, its dominant colours, the mean and standard deviation of image’s RGB values, the frequency of change in RGB values, the number of defined areas (“pixel groups”) enclosed within a well-defined boundary, the compactness of the principal (i.e. largest) pixel group, the major and minor axis of the principal pixel group, its circularity and its perimeter.

The challenge was – and remains – to select image features and a distance function such that the resultant distance really is a measure of image similarity: ideally the distance between the images, gauged on this metric would correspond to our subjective assessments of image similiarity.
Measuring the distance between two images which we judge to be alike would result in a relatively low aggregate value, while measuring the distance between two images which we judge to be quite different would result in a larger aggregate value.

How do you use these measures of similarity to help you decide what effects you will apply to images?

B: First, I wanted my program to emulate the filmmaking methods to which I have become accustomed. To this end, I formulated some loose rules that would capture some of my experience in deciding what image processing algorithms might be appropriate to images
that possess a given set of features. (Examples of such rules are: if there are a large number of pixel groups in the image and there are many changes in colour between adjacent pixels, then sharpening the image is not highly recommended; if the image is of very low contrast, then reducing the intensity of the image is seldom valuable; if the average size of pixel
groups is large, then applying algorithms that enhance the texture of the
image is a less valuable choice.) I created a program that employed a constrained random process -- the constraints based on these rules as well as on the image’s signature -- to decide which image processing algorithm or algorithms would be applied

to images.

The program looks at images and assesses their features, and based on what it discovers, decides which processing procedures most likely suit the image, and what procedures will be less likely (and how much less likely). Different features of an image are assigned different weights, and those features that are assigned greater weight are given a greater role in deciding which image
processing methods are desirable or undesirable (and how much less desirable or undesirable).
The application then chooses, by chance operations, a set of processing methods to apply to the database images.

Where would you like to take your work with this computer programme? How do you want to improve it?

B: I want to introduce better means for modelling a film- or video-maker's working methods, for capturing a filmmaker's (or videomaker's) understanding of what characteristics of the image make certain image-processing appropriate and other's inappropriate. The way I modelled
one's estimation of the appropriateness of a particular method to a given image was far, far too simple. What I did was simply to imbed in the program a "seat-of-the-pants" "guess-timate" of how undesirable a certain feature made a particular algorithm. For example, having a certain property might make using given image-processing methods either "slightly undesirable," or
"moderately undesirable," or "very undesirable" (each represented by a different weight), and more precise measures of a filmmaker's sense of the appropriateness of a method need be introduced.

I also incorporated a kludgy sort of "fail-safe" provision into the application. After applying the constraints I have described, the program selected one or more image processing methods to apply to the image, processed the image and displayed the result. The user was then asked to confirm that what he or she sees is satisfactory -- thus, instead of modelling the film- or video-maker's knowledge, I simply called upon it (and used it interactively). If the result was deemed satisfactory, the program applied a similar treatment to a set of similar images and saved the result to film.

All this needs to be drastically reworked. My "fail-safe" method of allowing the operator to interact with the program conflicts with my goal of refusing immediate authorial imposition. Further, I need to develop means to capture the "fuzzy logic" involved in these decisions. This could be done by building a learning component into the program that would enable the program to correlate the features an image possesses with the image-processing methods a particular film- or video-maker finds appropriate. Further, to make the program more flexible and better able to accommodate different ways of working, the user should be given the choice as to which sets of features, from a broader array of features than I now employ, would be relevant to determining which image-processing methods might be applied to the image.

Introducing fuzzy and neural learning into this application would have this benefit as well: the assumption that there can be standardised metric that corresponds to all users

judgements of image similarity is a doubtful one -- just as it is doubtful that all film or video editors take into account the same set of features when they are creating "plastic" cuts (edits based,
essentially, on the similarity of images), or even that an individual editor takes the same features into account on all occasions. Creating a system that would adapt to individual users (and, perhaps, even to individual circumstance) by being "re-trained" could allow for these variations.

Despite its current limitations, however, I believe the programme is a novel way of using image processing in film and video production. I also believe that the Cagean compositional ideas on which this application is based are rich and this makes me eager to continue to develop the project.

Project Description for a subsequent grant proposal to the Natural Sciences and Engineering Research Council of Canada

Over the past few years, I has been developing a computer application that has allowed him to use measures of similarity to constrain random processes that are used to decide which image-processing methods ("effects") should be applied to frames in a film or video. Bruce has worked on this strictly as a non-professional programmer, building a tool for his own use; he used this application in making his last two films. He was pleased with the results he obtained and, though he developed the application for himself, he feels that it might be of interest to other film- and video-makers. Accordingly, he wants to take this work further. In what follows we explain briefly his initial reasons for developing this application, give a summary of the evolution of the project, outline the enhancements we would like to introduce and discuss the technical and engineering challenges the further development of project poses. In our statement of the challenges involved, we explain the importance of developing this project collaboratively.

Stochastic techniques have a long history in music composition; but over the last fifty years they have become much more common. The richness of John Cage's writings on compositional method, and the resolute challenge they posed to conventional ideas about artmaking, inspired many composers. Cage protested against the idea that an artwork is the product of an artist's feelings – he found the notion that feelings should be allowed to dictate the ultimate form of the work anathema. He believed that the creative process should imitate nature in its manner of operation and strived to find creative methods that would accord nature a role in shaping the work. For Cage, chance operations figure among the methods that achieve this goal of selfless making. The richness of Cage's writing helped make the use of aleatory techniques common among composers. The rigour of writings by Iannis Xenakis and James Tenney – composers who, like Cage, took an interest in stochastic methods – and the power of their works re-enforced this influence.

Myproject has explored the possibility of extending these composers' ideas to the visual domain. The initial framework for this exploration was drawn from the James Tenney. Tenney made extensive use of measures of similarity in the analysis of music structures in his book *Meta+Hodos*, and subsequent composers applied those methods to generating series of musical events. I was intrigued by the possibility of developing analogous compositional procedures for working with sets of images and, in particular, by the possibility of using measures of similarity to constrain random processes. He decided to develop a computer application that would allow him to do this and, at the same time, would be consistent with the principles he used in composing films. He began to construct a tool that would emulate his way of working and would extend it, by eliminating subjective whim. He conceived this program as a means that would allow him to collaborate with the machine to produce "visual compositions."

Bruce first developed a rudimentary application that stored a set of images (that might constitute a number of sequences in a film) in a database along with a set of image descriptors ("meta-data") and a set of image processing algorithms. The application applied image-processing methods to the images in the database; the methods to be applied were selected by random processes that operated under the constraint of estimations of the similarities between images. Images were partitioned in groups based on the similarities indicated by their descriptors, as were the image processing methods (the decision on which methods most closely resembled other methods was completely informal and subjective) and the image processing methods to be applied to a reference image were chosen at random; the user was asked to confirm the choice and, after that, the methods most similar to the randomly chosen method were applied to the images in the database that most resembled the reference image. The scope accorded to randomness in the selection of the processing methods varied with the way the system is trained: when the reference images that were used to train the system had little similarity to the target images, the program relied more on random selection; when the reference images bore a strong resemblance to the target images, random selection had a less important role (and the use of methods that can take on varying attributes, and change with the degree of difference between the reference image and target images, was favoured). Processes that adapted to degree of difference between the reference image and the target images were used to create the impression of a continuous, ongoing change that sweeps across a set of images (that could, for example, represent a shot in a film or video).

Bruce used this application in a film (titled *Crack, Brutal Grief*) that he completed three years ago. This way of using image processing methods in film/video production interested him enough (and, he thought, the results were good enough) that he wanted to carry work on this application further.

The first improvement to make was obvious: using image descriptors was awkward and introduced an unnecessarily subjective element. He realized that the application would more truly reflect the compositional ideals to which he aspired (especially the Cagean ideal of avoiding authorial imposition) if he were to make use of methods to "compute" the similarity between the two images algorithmically.

Though in his first blush of enthusiasm for the project he did not know it (he did find out fairly quickly), methods for computing the degree of similarity images have to one another were already in widespread use. Indeed, ways to search a database for images that resemble a "key image" (or "query image") have become relatively well understood. Typically this process takes place in a number of steps:

Load the key image or query image (the image for which we want to find similar images).

Utilizing methods of feature extraction, measure a number of features of the key image. This stage creates a “signature” for the image.

Load each image in the database and generate a signature for it.

Calculate the Euclidean distance between the signature for the key image and the signatures for each of the database images. Sort and store these values – what results is a list that shows the proximity (based on its signature) of each of the database images to the query image.

Bruce worked out a version of this basic framework for himself. The features he used to create a signature for an image were its intensity, the dominant colours of the image, the mean and standard deviation of its RGB values, the frequency of change in RGB values, the number of areas (“pixel groups”) enclosed within a well-defined boundary, the compactness of the principal (i.e., largest) pixel group, the major and minor axis of the principal pixel group, its circularity and its perimeter.
Thus, the resemblance between images was measured by the Euclidean distance between points in a multidimensional feature space. The challenge was – and remains – to select image features and a distance function such that the resultant distance really is a measure of image similarity. Ideally (though this is not entirely practical in reality), the distance between the images, gauged on this metric, would correspond to our subjective assessments of image similarity – thus, measuring the distance between two images that we judge to be alike would result in a relatively low aggregate value.

The second improvement was to incorporate a very rudimentary expert system into the application. (An expert a computer application that performs a task that would otherwise be performed by a human expert; often the expert’s knowledge is captured in a set of rules.) Because Bruce wanted the program to emulate and to extend his customary working methods, once he had constructed a program for identifying and measuring key features of the image, he formulated some loose rules that would capture some portion of his experience in deciding what image processing algorithms might be appropriate to images that possess a given set of features. (Examples of such rules are: if there are a large number of pixel groups in the image and there are many changes in colour between adjacent pixels, then sharpening the image is not highly recommended; if the image is of very low contrast, then reducing the intensity of the image is seldom valuable.) Bruce created a program that employed a constrained random process – the constraints were based on these rules as well as on the image’s signature – to decide which image processing algorithm or algorithms would be applied to images.

The program would look at an image, assess its features and, based on what it discovered, would decide which processing procedures would most likely suit the image, and which would be less likely (and how much less likely); different features of an image were assigned different weights, and those features that were assigned greater weight were given a greater role in deciding which image processing methods to use. The application would then chose, by a stochastic process (constrained by this system of weights), a set of processing methods to apply to the database images.

Bruce has made parts of two feature-length films (*Crack, Brutal Grief* and *Eros and Wonder*) using this software. He believes that he has developed a novel way of using image processing in film and video production. He also believes that the Cagean compositional ideas on which this program is based have great potential for being extended into the visual domain – far more potential than anyone has yet explored.

The enhancements I would like to introduce can be divided into three categories. First, we want to bring aspects of the project into conformity with recognized methods in image analysis and retrieval. Second, we want to improve the ways the application learns about a film- or video-maker's working methods. And third, we want to expand the application image processing capabilities. The technical methods we are proposing to address these issues are detailed as follows:

1. Image Analysis and Retrieval

There are well-established methods for measuring image similarity and for image analysis;these methods can be adapted to analyzing the sets of images in order to select appropriate image processing methods. We intend to introduce more standard methods for gauging image similarity than Bruce used in his version of this application. This will require taking into account several sets of features.

One set of features commonly used to generate an image's signature depends on transforming the image into the frequency domain. We intend to introduce features that incorporate the frequency components of the image into our calculation of the image's signature.

The second set of features will be derived from modeling wavelet coefficients by means of a Mixture of Gaussians (MOG) model. (This requires the use of wavelet transforms: transforms are mathematical techniques that are applied to signals to reveal information not apparent in the raw signal. Wavelet transforms show us what frequency bands exist in a signal at what time intervals – that is, they give us information about time and frequency simultaneously, while raw signals generally furnish only time-domain information.) Wavelet coefficients have been proved to be effective at characterizing signals at multiscales. However, one set of coefficients for each scale of wavelet transform is not sufficient to characterize complex scenes. The reason is that, as a high frequency detector, the wavelet coefficients have a peaky, heavy-tailed marginal distribution. A small portion of coefficients takes large values – this phenomenon marks where edges and/or textures occur – while most other coefficients take small values. Therefore, we separate the coefficients into the edge component, the texture component and the smooth component, in each scale of the transform, using a three class MOG to model them in each scale. The determination of the coefficients in the three classes and the calculation of the statistics are achieved by a two step process:

Separating the coefficients into the three classes. The separation of the smooth areas from the edge-texture areas is relatively simple: it can be done using local statistics such as pixels' variance. Further separation of edges from texture regions proves to be a much bigger challenge due to the similarity in the simple statistics of the two groups of image regions. We propose applying the recently proposed Edge-Texture Characterization (ETC) measure to this task. By calculating the degree of gray level correlation within a local image region, the textureness/edgeness of the region is statistically characterized, with a higher correlation for textures, and a lower correlation for edges. Because it is difficult to define the concepts of “edge” and “texture” for the ETC in terms of crisp sets, fuzzy set theory is applied for characterizing these concepts. (Fuzzy sets are distinguished from normal, “crisp” sets in that any element is either a member, or a non-member, of a normal set – in a fuzzy set, on the other hand, an element can be member of the set to some degree, and a non-member to some other degree.)

After the three coefficient classes for each scale are identified, the MOG models are applied to characterize the distribution of the coefficients and their corresponding statistics. The well-known Expectation-Maximization (EM) algorithm is used to obtain these coefficients. (The EM algorithm is an algorithm that attempts to estimate complete data from incomplete data. It does so by repeatedly estimating the “likelihood” function and finding the set of parameters that maximizes the function. The method has application in image segmentation.)

The third set of features we shall take into account is the family of features required to perform object-based analysis and to tracking regions of interest (as detailed in Section 3 “Expanding the Capability of the Application”). This set of features is extracted using more vigorous image processing techniques. The approach we propose is to first apply the novel perception-inspired segmentation algorithm based on a hierarchical cluster model (HCM) and a self-organizing tree map (SOTM). The HCM and the SOTM are both neural networks. (A neural network is a technique used in artificial intelligence that imitates the way a human brain works; it works by creating connections between processing elements, which serve as the computer equivalent of neurons. The organization and weights of the connections determine the output.) The SOTM is a novel tree-structured learning architecture for unsupervised data clustering. (Unsupervised methods are mathematical methods that, without a teacher or supervisor, try to identify internal structures of a given dataset and group elements that bear some similarities with one another.) It has been demonstrated that the SOTM is more capable than the well-known SOM (“self-organizing map”) architecture of clustering data that are spread evenly and thinly over large parts of the image and are difficult to associate with a particular object. The HCM is a special type of neural network that attempts to mimic the human visual cortex. It is a multilevel neural network – a series of interacting neural networks in which the state of one network affects the connections in another network; thus, it consists of hierarchically linked sets of neurons.

When performing image segmentation, we first use the SOTM to cluster pixel groups: regions of connected pixels are formed, based on homogeneity. Using values that represent visual features whose importance is emphasized in Gestalt psychology as the states of the neurons in the network, the HCM groups the regions of connected pixels into semantically and visually meaningful objects. We are then able to extract the salient spots, and extract the salient features, from the segmented image.

Since human users are the ultimate judges for ranking the similarity between visual files, it seems intuitive to incorporate perceptual knowledge into the ranking process. Although many ranking algorithms exist, they are either linear or quadratic in nature, and cannot properly capture the non-linear characteristics of human perception. Therefore, we propose to use a radial-basis function network, a type of “feed-forward” network that uses radial functions as the kernels. (Radial functions are a class of functions whose characteristic feature is that their response decreases, or increases, monotonically with distance from a central point – so an RBF network can be described as constructing global approximations to functions using combinations of basis functions centered around weight vectors to simulate the perceptually inspired similarity ranking process.) It has been shown that such a model is capable of capturing important characteristics of human perceptual knowledge in image retrieval. Because the model has a built-in normalization capability, the tedious normalization process associated with the linear ranking algorithms is eliminated.

Based on the work described above, we create an expert system that will decide which methods might be applied to which images, by examining their features. Features that might affect the methods that apply will probably include wavelet features as well as more conventional features such as smoothness, texture, edge, contrast level etc. When evaluating the influence that its features have on determining which processing methods are suitable to an image, the combination of features should be evaluated on a numerical scale (the image possesses feature A in degree X, feature B in degree Y, and feature C in degree Z, making method Q appropriate to degree M) rather than by logical ruth-table-like framework (if the image possesses feature A and feature B, but not feature C, then method Q is appropriate); this makes a fuzzy neural network (a fuzzy neural network is a neural network whose kernel functions belong to fuzzy sets) the ideal tool to implement the process. Accordingly, we intend to make use of this technology.

2. Adaptive Learning: Capturing Filmmakers Working Methods

The second category of improvements, that comprises the most important improvements we wish to introduce, have to do with modeling film- or video-maker's working methods. We propose to find better means for capturing a film-maker's (or video-maker's) intuitive understanding of which characteristics of an image make certain image-processing methods appropriate and others inappropriate. Elder's way of modeling a filmmaker's estimation of the appropriateness of a particular method to a given image was elementary. What he did was simply to embed in the program a "seat-of-the-pants" "guess-timate" of how undesirable a certain feature made a particular algorithm: having a certain property might make using a given image-processing method either "slightly undesirable," or "moderately undesirable," or "very undesirable" (each represented by a different weight). More precise measures of a filmmaker's sense of the appropriateness of a method need be introduced (by using machine-learning). Moreover, to simplify the programming, Bruce constructed the program as though a given image either has a given property or lacks it completely: he did not take into account the fact that an image may have a given property in greater or lesser degree. Furthermore, he incorporated a kludgy sort of "fail-safe" provision into the application. After applying the constraints described, the program selected one or more image processing methods to apply to the image (the number of methods applied to a particular set of images could be selected by the user), processed the image and displayed the result. The user was then asked to confirm that what he or she sees is satisfactory – thus, instead of modeling the film- or video-maker knowledge, Bruce simply called upon it (and used it interactively). If the result was deemed satisfactory, the program applied a similar treatment to a set of similar images and saved the result to an appropriate medium. (Bruce prefers to output the result to film, but video might serve some others just as well.)

These earlier efforts at modeling a film- or video-maker's knowledge needs drastic revision. For one thing, the "fail-safe" method of allowing the operator to interact with the program conflicts with the goal of refusing immediate authorial imposition. The means of capturing a film- or videomaker's knowledge has to be reworked from the ground up. In fact, I see my chief role in the proposed project as steering the formulation of the requirements specifications associated with modeling the creative processes that video-makers and film-makers use (or, to make a less sweeping claim, that he has sometimes used). To do this, we need to take into account, first, that an image's having a certain property in greater (or lesser) degree makes certain image processing method more (or less) suitable – we need to develop means to capture the "fuzzy logic" of this situation, so typical of the creative process. Furthermore we should take a more systematic approach to gleaning a film-maker's intuitive knowledge. This will be done by building a learning component into the program (using fuzzy neural learning) that would enable the system to correlate the features an image possesses with the image-processing methods a particular film- or video-maker finds appropriate

Furthermore, the manner used to indicate the appropriateness, under given conditions, of certain processing methods needs to be reworked from the ground up. Concepts like "slightly desirable," "moderately desirable" and "very desirable" can be better implemented using a hierarchical modular fuzzy neural system, and specifically a class-in-expert hierarchical structure (CEHS). At the grand level (the expert level), each concept is modelled by a network module (an expert). The appropriateness of an image processing method to an image is measured by the composite input from the experts, determining the degree to which this method fits into the concept models. At the finer level (the class level), the concepts are developed based on such image characteristics (classes) as its smoothness, texture, contrast, etc., and on knowledge of the association film- or video-makers make between these features and appropriateness of specific image processing methods.

Introducing fuzzy and neural learning into this application would have this benefit as well: the assumption that there can be standardized metric that corresponds to all users' judgments of image similarity is a doubtful one. Creating a system that will adapt to individual users (and, perhaps, even to particular circumstances) by being "re-trained" could allow for these variations. Although it is unlikely that in the foreseeable future anyone will be able to create a system that comprehends all factors that might influence decisions about the suitability of particular processing methods and takes into account all the idiosyncracies that, in a given circumstance, might influence a particular individual's choice of processing methods, categorizing these factors, and modeling a limited number of these categorizations, using a CEHS with higher level hierarchy, is certainly a realistic goal, and achievable within the scope of this project.

To make the program more flexible and better able to accommodate different ways of working, the user – actually, this provision requires a knowledgeable user – should be given the choice which sets of features, from a broader array of features than we now employ, would be relevant to determining which sequence of image-processing methods might be applied to the image. We would like to have the system facilitate such choices. Creating such correlations between image features and processing methods is a difficult task: it involves a complex multivariate analysis of features. Though optimal feature selection is complicated, there do exist simple sub-optimal solutions which can effectively perform multivariate analysis of features. One such method is based on feature study by sequential feature selection (SFS), and effective ranking by general recurrent neural networks (GRNN). SFS is a search method in the space of feature subsets that generate the subsets incrementally, in increasing or decreasing groups, either beginning with the one feature that maximizes performance for a group of one and adding another that maximizes performance for a group of two, etc., or, beginning with a full set of features and gradually removing that feature whose removal yields maximal performance improvement. In the context of the work on hand, the target output is the effect a particular sequence of image processing methods produces, and the input is a subset of the complete feature set. By incrementally or decrementally changing the number of features in the input subset, we identify a set of features that can cause a given effect. Without tedious learning, the GRNN can determine the effect of features in an efficient and timely manner. These enhancements provide a far more elegant way of determining the suitability of different processing methods to the image and, moreover, better preserves those Cagean ideals ideals relating to refusing authorial imposition that we are striving to embody in the application, than the use of rules relating image features to processing methods currently employed.

3. Expanding the Capability of the Image Processing Application

The final category of improvement we propose to introduce is to expand the range of image processing methods that we apply to the image. In expanding the capabilities of the system, we shall emphasize the use of regions of interest (ROIs) and of geometric transformations. (Bruce did target regions of interest in some of the methods employed in earlier versions of the program, but he did not carry this far enough.) The expanded use of regions of interest is crucial to emulate a video- or film-makers' customary working methods: when one is "composing" a film/video, particular regions of the image often assume greater importance, though, of course, there is a balance of interests between the focus on a particular area and the concern with overall structure of the entire image. We intend to introduce means that allow the application to accommodate sub-image queries and queries that give a higher weight to features in a specified area of an image but also take into account the image's overall character. To start with, we will apply the perceptually inspired image segmentation algorithm described in Section 1 Image Analysis and Retrieval to locate the ROIs.

At present we employ only a limited number of means for creating geometric transformations; we would like to develop a rich set of devices for modifying the geometry of the images. Such means are especially valuable for creating a sense of "forward motion" in an effect – the impression that an effect sweeps across the frames that make up a shot. In previous versions of the program, Bruce relied on effects that can take on varying attributes: frames which came later in the shot (and so, generally, were more different from the first frame, which served as a reference image) were treated differently that frames that came earlier – the parameter(s) which controlled the varying feature(s) changed with the degree of difference between the reference image and target images. Though these rudimentary methods did work (in the sense that Bruce was able to create film sequences using them), they are limited. We propose to create a richer set of means for engendering a sense of "forward motion" than those we currently employ, by using an improved version of the image-similarity metric that can examine how far the pixels in certain region of the target image have migrated from their location in the reference image and by employing that measure to influence the amount of "distortion" the images undergo – the greater the distance a pixel group has travelled, the greater the degree of transformation.

There are motion estimation and compensation techniques (established in video processing and computer vision) that allow one to determine by how much the location of a pixel group in a particular images differs from its place in a reference image. By grouping the connected macro-blocks which have similar motion vectors, we can estimate the motion of the region of interest to a specified degree of accuracy. We propose to adopt the Particle Filtering algorithm in tracking down the ROIs after they are extracted by the object segmentation algorithm. Particle filtering techniques are tracking methods based on statistical mathematical routines, and were developed to address the problem of tracking contour outlines through heavy image clutter. The filter output at a given time-step, rather than being a single estimate of position and covariance as in a Kalman filter, is an approximation of an entire probability distribution of the likely positions of the region being tracked. This allows the filter to maintain multiple hypotheses and thus to be robust, even when it confronts distracting clutter. Such measurements could be used to determine the degree to which the image's geometry would be altered. A series of images (for example, the images making up a shot) would exhibition progressive transformation. This would be an effective way of creating a sense of forward motion. Introducing this capacity interests us especially – we believe that this capacity would make the program of interest to many film- and video-makers.

The analysis of motion will be extended to measures of motion similarity, i.e., measures indicating how closely the profiles of two movements match one another. Video indexing techniques can be applied to the estimation of motion similarity. Instead of comparing key frames, as most video indexing systems do, this method studies the dynamic nature of video sequences via a template-frequency model to capture the time-varying nature of videos. The computational measurement of melodic similarity in music is lively field: melodic similarity may be established by comparing profiles of pitch contours of melodic lines (think of line that rises when the pitch rises and that falls with the pitch falls); these techniques play an important role in some approaches to algorithmic composition. We believe that measures of movement similarity could play as important a role in our approach to film- and video processing: the similarity between the contours of two motions can be gauged and that measure could be used as a constraint on the selection of image processing methods.

Among the potentials for further exploration are the use of constrained random processing in sequencing images and the use of aleatory methods for interactively selecting and displaying images from the database. However, these remain topics for further investigation and, strictly speaking, are not part of this project.