

Audiovisual Poem : Synthesis of Image and Music

Nakamura, Shigenobu

Department of Contents and Creative Design, Graduate School of Design, Kyushu University

<https://doi.org/10.15017/18916>

出版情報 : 芸術工学研究. 13, pp.47-56, 2010-12-15. 九州大学大学院芸術工学研究院
バージョン :
権利関係 :



Audiovisual Poem:
Synthesis of Image and Music

映像音響詩

－映像と音楽の統合－

NAKAMURA Shigenobu

中村滋延

著者抄録

本稿は私が 1990 年代半ばからこれまでに制作した映像音響詩 13 作品についての解説である。様式から言えば、映像音響詩は音による表現を第一義に考えてつくられている「映像作品」である。しかし表現の目的から言えば、映像音響詩は視覚的要素を構成に取り入れた「音楽作品」である。そのことで音の意味生成の可能性を豊かにし、音による想像を刺激し、音だけではなし得なかった実験的な音楽表現を様々に試みている。その試みの紹介を中心に、個々の作品において作者が表現に対して何を目指したのかを記述したものである。

本稿では最初に映像音響詩の概念を述べ、13 作品を制作年順に、ただし作品を「映像と音響のモンタージュ」(*Walk, Epitaph*)、「音楽形式の応用」(*Yin-Yang*)、「評論としての映像音響詩」(*sabi, Metamorphosis of Love*)、「抽象型映像音響詩」(*Life, Common Tragedies in Urban Life*)、「引用型映像音響詩」(*Play, Lust*)、「個人映画としての映像音響詩」(*Naga Variations*)、「動く造形美術としての映像音響詩」(*Mandala Fantasy, Impatience*)、「物語映画としての映像音響詩」(*Samsara*)に分類して、解説している。

What is an Audiovisual Poem?

"Audiovisual poem" is a phrase I coined myself. (In German, the phrase is audiovisuelles Gedicht, and in French, it is poème sonore visuelle.) It is a visual work created with a focus on using sound to reflect the true essence of the composer. The function of visual expression in an audiovisual poem is to deepen the expression of sound and to expand on it. I might dare to say without fear of misinterpretation that there is no essential difference for me between the composition of a conventional musical work and the production of an audiovisual work.

I have been involved in the production and presentation of music theater since the 1980s, viewing it as a metaphoric pillar of my composition work. The music theater features musical compositions to which visual elements are added through stage performance (players' movements, characters' gestures, stage art and set lights, etc.). In the production of an audiovisual poem, the visual element seen on stage during a musical theater production is replaced with an image on a television monitor or screen.

The following is a brief look at the process of integrating a visual element into a musical composition.

Composers can invent music not only by using individual sounds but also by combining two or more interrelated sounds. It is not until the sounds are heard in relation to other sounds that they gain their full meaning. This meaning is of a structural nature, like tension, stability, slack, and reservation, for example. With visual information, the structural meaning of the relationships found in the sounds expands; the possibilities are amplified or converted, changed to a negative form,

and so on.

Let us look at an example of a "C" note played softly on the piano. Consider a scenario where the preceding, following, or even simultaneous action has some bearing on the played note. If, for instance, the pianist raises his or her arm up high and proceeds to strike the keyboard with great force, the audience will naturally expect the note to ring out with the equal force. However, if the note is in fact softly played, there arises a feeling of unexpectedness, a sense of evasiveness. Such a feeling can only be instigated with a visual impression. Simply put, visual impression gives further meaning to sound.

Visual expression in the audiovisual poem takes on that exact role. Depending on the structure of the piece, the impact produced by the visual component often exceeds the impact of the musical composition itself (structural meaning or concrete meaning as indicated by images and movement). Thus, at such a juncture, various possibilities arise between the visual and aural elements. Such a juncture is the starting point of my audiovisual poem.

Walk: Montage of the visual and the audio(1)

Walk, a 10-minute piece produced in November 1994, is the first work I clearly identified as an audiovisual poem. The main audio components are footsteps. Voices and urban clutter are also used as sound components. I recorded every component, and then edited them using a computer. The audio part of the piece is composed mainly using the sounds of footsteps, creating differentiation in the strength, tone, pitch, range, rhythm, and cluster density of the sounds. This is called "absolute music" which consists of the sound components of footsteps. The main visual components are figures of legs, feet, and people walking around. The relationship between the musical and visual parts is as follows.

In the opening, the sounds of footsteps are played repeatedly at certain intervals with a variety of sound reverberation effects. These sounds of footsteps coincide with the image of feet (Fig.1). The image of the feet cutting into varied sizes comes sometimes with the sounds, or sometimes ahead of or behind the sounds. This gives the audience different images or musical meanings even though almost the same sounds of footsteps are repeatedly played.

At the end, the sounds of footsteps are modulated like

humans calls and repeatedly played at intervals. With the sounds, the image of a yellow quadrangle appears and disappears (Fig.2). Then, by slow degrees, the image of the quadrangle comes behind the repetitive sounds. Soon, the appearance of sounds and images are completely out of alignment. The audience can hear the sound in their imaginations every time the image appears, even though no sounds are played. This is because the image becomes synchronized with the sound of footsteps, and eventually the new rhythm could be created with actual sounds and imaginative sounds.

This piece is highly appreciated as "music." It was introduced as the best Japanese computer music in the *Leonard Music Journal*, Volume 5 1995 (MIT Press, Cambridge MA, 1995) and the audio part was recorded to the accompanying CD. However, in terms of the visual piece, I have faced a few challenges. One of them is that the composition of the visual part is totally subject to the audio part, which means the visual part cannot be independent. The other is that, once a concrete figure appears in the visual part, the figure itself starts to tell the story to the audience because its concrete meaning evokes the imagination. That makes the audience try to read the story by seeing the images of feet as the leading motifs; however, the audience becomes confused because the piece does not have a story.

It is possible to insist that this piece is just music with the visual part as a complement, but it would be unfortunate for attractive sides of the visual part would be abandoned.

Yin-Yang: The visual component as the application of the musical form

In *Yin-Yang*, an 11-minute piece produced in April 1995, I tried to conquer the challenges I faced in my previous work, *Walk*. These were to make the visual part independent (the literal meaning of "composition"), and to make the audience naturally understand the piece even when the visual motif itself starts to tell the story.

To make the visual part an independent composition, I tried to match the visual motifs to the musical form, using an "introduction," "exposition of theme A," "exposition of theme B," "development," "reprise of theme B," "reprise of theme A," and "coda." For example, the motif of theme A is a dark shadow of a large knot of rope cutting across the center of the

screen, which sometimes can be seen as the shadow of animals or the shadow of a human being hanging on to the rope (Fig. 3). The motif of theme B is a dark shadow of palms covering almost the entire screen (Fig.4). Each motif emerges slowly by itself during the exposition period. In the development period, one motif can work with other motifs as a background, and it can even move at a different tempo from the exposition period. This means that a motif can vary from itself. In the reprise period, a motif appears slowly, not by itself, but with other motifs. Theme A and theme B come in no particular order.

No specific story is presented in this piece, but the motifs are prepared and composed to make it easy for the audience to feel the story and build up the story by themselves. The motif of theme A captures humans' religious distress, and the motif of theme B captures humans' unlikable lust, especially sexual desire. The motifs relating to these themes were selected to portray contrast, as follows: woods in the shrine representing the holy world, and a pictorial offering representing the secular world; cherry blossoms in full bloom representing light, and sewage water puddles representing dark; an adult female figure representing prostitution, and the figure of a girl representing virginity. Each contrast is intended to emphasize the meanings of the motifs.

I faced two other challenges after completing the work. One is that the method of signification, implication, and representation is too arbitrary. This leads the audience to have difficulty finding the story through the visual part. The other is that I high-handedly applied the musical form to the visual part. Because the visual part is not necessarily composed to match up with the audio part, the audience might feel that the visual part is sometimes inconsistent and broken.

The audio part consists of electronic modulated flute solo sounds and other various electronic sounds based on concrete sounds. The final sounds are related to the motifs of the visual part, though the meanings of the sounds are not significant. The music is basically absolute music. The concrete sounds are positioned as accompaniment to the flute solo. The audience simply accepts the visual part as the image of the audio part.

Fortunately, this piece was nominated for an *International Video Art Award* (now called the *International Media*

Art Awards) in 1995, at a presentation hosted by SWF South-east German Broadcasting Co. and ZKM Media Art Center. It was aired in Germany, Austria, and Switzerland. This gave me confidence that my video art could be accepted widely by the public, and made me work harder to produce another series of video art pieces.

Epitaph: Montage of the visual and the audio(2)

My principle for producing an audiovisual poem is to invent simultaneously both audio and visual parts. In actuality, however, the process depends on the contents of the works. I sometimes compose the audio part earlier and then produce the visual part, or vice versa. *Epitaph*, a nine-minute piece produced in May 1995, embodies my principle of making the audio and visual parts at the same time.

Half a year after my father's death, I found some home movies of my father. I had never expected to find him in the videos, because he always video-taped our family by himself. I did not shed any tears when he died, but I could not stop crying when I saw my father in the video. This made me realize the power of video. This is when I decided to produce the requiem for my father as an audiovisual poem. I finished producing the piece in just one week with the main motifs of my father's pictures and voices.

The motivation of this work is dependant on my personal emotions. Therefore, the emotional expression of the piece could have been too overabundant to maintain the consistency of the work. Sometimes, a piece is rather subjective for the audience to understand the work deeply. Therefore, for this work, I made composing in an objective manner, which is my top priority. Then, I limited the number of motifs. The main motifs of the visual part are just two short shots of my father (Fig. 5, Fig.6), and the main motifs of the audio part are just two phrases spoken by my father: "Arigato" (thank you), and "kocchi ni misete" (show me, please). These motifs are repeated often, and they are changed repeatedly. For example, the cycle of the repetition is jumbled, and the synchronization of the parts is uncertain. In addition, the main motifs themselves are transformed or modulated by replacing or reduplicating other motifs. The transformation and modulation as such is "variation" in musical composition theory.

In the introduction, abstract figures appear in the visual

part, representing tracking back the memories. The sounds are played in response to the visual part. In the coda, a gravestone emerges in the visual part to represent the repose for souls, and a sutra is heard in the audio part. These are the measures taken for the audience to be able to understand the story in the main motifs.

This piece was nominated for an *International Video Award* in 1996 and aired in Germany, Austria, and Switzerland. It was also accepted in the *l'immagine leggera* video art festival competition in 1997 (Palermo, Italy) and performed there. These achievements made it possible for me to work as a visiting artist at the ZKM Media Art Center in Germany from 1997 to 1998.

sabi: The audiovisual poem as a criticism (1)

As a visiting artist at the ZKM Media Art Center in Germany, I produced *sabi*, an 11-minute audiovisual poem, in January 1998. *sabi* is a word taken from the Japanese phrase *Wabi Sabi*, which can be translated as "austere refinement and quiet simplicity."

The main musical components of this piece are sounds of *Gidayu*, the background music of *Bunraku*, a Japanese type of puppet play, and the main visual components are modulated video images of me living in Karlsruhe, Germany, where ZKM is located. This piece was produced based on three basic ideas, which are sometimes nested in the piece. The first of the three ideas was to compose the music in the form of *musique concrète*. The music represents the scenario of *Gidayu*: *Kumagai Jinya no Dan*. The second is to present the ideas through visual images both symbolically and abstractly (Fig.7). The third is to picture a visiting artist, me, living in a foreign country with considering or bothering, with the music and visual images (Fig.8).

While in Karlsruhe, I often went to the opera and was overwhelmed with excitement. It made me wonder what had been created in Japan when opera was active in the west. During this time, by chance I listened to a CD of *Gidayu* and was impressed by its very sensitive expressions. Considering *Gidayu* to be equivalent to the opera, I listened to it carefully every day. Then, I produced this work in order to present the attractive sides of *Gidayu* in my own way. In that sense, this work is a kind of criticism of *Gidayu* using the form of the

audiovisual poem.

Life: An abstract audiovisual poem (1)

Life, a six-minute piece produced in September 1998, was my first attempt at an abstract audiovisual poem. The piece consisted of nine parts: a, b, c1, d1, e1, c2, d2, e2, and c3. The parts a, b, c, d, and e symbolically represent birth, growth, chaos, cosmos, and loneliness, respectively.

These concepts are depicted in the visual part using the geometric figures as motifs (Fig.9 and Fig.10). By consistently using geometric figures as motifs, repeating the combination d-e-c, and inserting c at periodic intervals, the unity of the work's composition can be maintained. The variety of the work is also maintained through the variations that occur when c, d, and e appear, and are modulated musically.

The audio part is based on sounds that show the concepts of birth, growth, chaos, cosmos, and loneliness. For example, infants crying represents birth, and sounds of urban clutter represent chaos. In addition to these sounds, musical segments from Mozart's opera *The Magic Flute* are used, but these are extremely fragmented so that they are just materials for musical reconstruction. Nevertheless, I used the musical excerpts from *The Magic Flute* because the concepts I tried to picture in *Life* and the contents of the opera are deeply related to each other. *The Magic Flute* is a "hidden reference" in this piece.

Play: An allusion-type audiovisual poem (1)

Play, a 5.5-minute audiovisual poem produced in September 1998, was my first work categorized as an allusion-type audiovisual poem. It was originally designed to be an abstract audiovisual poem as a companion piece to *Life*. At the very beginning of my inspiration for this work, I symbolically depicted various scenes of ball playing as the components of the visual part. For example, I tried to present different types of ball playing, such as juggling, tennis, volleyball, soccer, baseball, golf, and bowling, with geometric figures. I tried to use these figures as the hints for the composition. I had no intention to make the audience understand the scenes of ball playing. What I intended was just to develop the abstract images and to continue showing them to the audience.

The audio part is absolute music even if it has human voices or concrete sounds as its components. Music components emphasize sound characteristics rather than significance. In relation to the visual part, my musical inspiration was based on the speeds of ball playing.

However, I found that geometric figures would bring scenes of ball playing to the audience's minds more than I expected. In order to avoid that, I applied motifs of photographed images in the geometric figures (Fig.11). That successfully made the images of ball playing back away. For example, in the scene that hints at soccer, collages of photographed female nudes are expressed as balls, and collages of photographed men's faces are expressed as soccer players (Fig.12). The photographed men's faces (players) strongly follow the photographed female nudes (balls). This makes it possible to portray humans' lust cynically. The significance of the photographed images overrides the images of ball playing.

Common Tragedies in Urban Life: An abstract audiovisual poem (2)

Common Tragedies in Urban Life, produced in April 2000, is a six-minute audiovisual poem in which the tragic side of urban life is cynically portrayed. The visual component is constructed using geometric shapes and their movements to symbolize the sadness brought about by various events in urban life (Fig.13).

At the beginning of the piece, 12 small circles, arranged on one big circle, blink violently while 12 variously colored points depict the modern city dweller battling against time. In the middle of the piece, some circles move hurriedly in a distorted lattice figure (Fig.14). The portrayal is of the circles confined in lattice, and of the lattices violently clashing into each other.

However, the meaning may be lost to the audience with the images alone being so abstract. The audio component thus adds to the meaning: the sound of a clock at the beginning of the piece; breathing, groaning, and panting; the crashing of various objects in the middle of the piece; the seeming chasing after time; the stress of urban life. In fact, such sounds are made not only to place an emphasis on the concrete meaning, but also to enhance and modulate the process so that it functions effectively as a musical composition. For instance, the

sound of the second hand is processed and modulated to express a sense of hurriedness.

Lust: An allusion-type audiovisual poem (2)

Lust, produced in April 2000, is a six-minute audiovisual poem. There were three aims in the planning of the piece in order to create an allusion-type audiovisual poem. The first was to compose the visual part from abstract shapes. Seven oblong shapes stand in a row on a black background. The lengths and widths of the oblong shapes constantly change, as they move right to left. With the consistent moving of the seven oblong shapes, variously changing yet always queuing, the unity is in the diversity.

The second aim was to make a photographic image into an abstract shape (Fig.15, Fig.16). A concrete message is applied to the abstract shape with images extracted from magazines, newspapers, and television, and with photographs taken from daily life. The message here is a warning regarding human lust and the actions that follow, resulting in crises in the real world.

The third aim was to make the sounds of the audio component complement the abstract visual components. The audio component is mainly composed of various concrete sounds and a piano. The concrete sounds integrate directly with the photograph part of the visual component, emphasizing the meaning of the life event. Then, as the piece continues, the concrete sound becomes spasmodic, without any connection to the background. It then is replaced with a piano, which then dominates.

I think the aims are achieved. While the audio part is so divided between concrete sound and piano music that it seems to simulate the soundtrack of an ordinary movie, it is preferable that the audio part of an audiovisual poem forms a whole that cannot be separated.

Metamorphosis of Love: The audiovisual poem as criticism (2)

Metamorphosis of Love, produced in March 2000, is an eight-minute allusion-type audiovisual poem. In this work, however, the allusion (or the quotation) is not just a means of expression, but is in fact the actual content itself. In this respect, it is different from the allusion in other works.

Here, I have attempted not only to include a number of Picasso's pieces that express love toward the model Marie-Therese, but also to analyze them and thus, to draw on Picasso's magnificence. It is a critique of Picasso expressed through images and music as they progress isochronally. The visual components' motifs are, therefore, all from Picasso's masterworks.

Picasso's paintings appear one after another as visual components. The order of appearance of the paintings begins with his encounter with Marie-Therese and continues to the end of their love, as indicated by the appearance of Dora Maar.

Picasso draws Marie-Therese not as a mere model, but as one with strong allegorical references to the maiden sacrificed to "the Minotaur," portrayed as Picasso himself. In order to express the skillfulness with which Picasso's pieces were composed, and what the piece was expressing, I processed the various paintings, blocking out certain parts, moving certain people, and changing motifs. Certain people and motifs are made to overlap, the color is changed, and two or more paintings are combined. While the critic discusses an artist or his work using words, I attempted here to comment on Picasso's works and love of Marie-Therese in the form of an audiovisual poem.

This work, however, is currently not available for screening to the public, because of the copyright infringements involved in using Picasso's work.

Naga Variations: The audiovisual poem as an individual film

Naga Variations, a six-minute audiovisual poem produced in March 2005, features my memories of trips to Angkor Wat in Cambodia. I categorize it as an individual film. I have been fascinated by Angkor Wat in Cambodia and have visited there several times. This piece is based on the pictures and videos I took on these trips, but I did not intend to make this work just a record of the trips. I did my best to make it an advanced art expression.

The components of the visual part are the Naga statues in the Angkor ruins, the landscapes in Siem Reap where Angkor Wat is located, with the visual "white noise" of TV monitors representing tracing back in the memories. The main

components are the statues of Naga, which appear on the screen modulated or varied (Fig. 17, Fig.18)

The main components in the audio part are human voices as concrete sounds, ethnic music, and sounds from Siem Reap and the Angkor ruins. Here, the significances and nuances of the concrete sounds are recognized as important. For example, the concrete sounds, such as the voices of playing children, the voices of tourists visiting Angkor Wat, noise in the market, and ethnic music being played on the street, are intertwined with the visual part by emphasizing their significances and nuances. The sounds of TV white noise and synthetic computer sounds are sometimes used to compensate for the concrete sounds.

Naga Variation was originally composed to show the motifs of the landscapes in Siem Reap and the Naga statues one after another, and these motifs were to be modulated or varied at every appearance. However, presenting them alternately made this piece too schematic, so I deleted many of the daily landscapes in Siem Reap for the completion of the work. The variation of the Naga statues is the main attraction of this work.

Mandala Fantasy: The audiovisual poem as moving fine art (1)

Mandala Fantasy, a five-minute audiovisual poem produced in March 2005, presents the enigmatic attraction of Mandala using traditional Mandala as well as Mandala that the author himself has created. Mandala is attractive enough as a still image, though more attraction can be brought out by picturing it as moving fine art.

The visual part of this work is a kind of animation. Most of it uses the animation technique of moving fragmented pictures as its characters in two dimensions, though some of it uses photographic images. In the animation, Mandala is continuously developed, such as Mandala made of candles, Mandala of Brāhmī script, Mandala of materialized human faces or bodies (Fig. 19), new Mandala made of decorated human faces or bodies, new Mandala cut off in a circle (Fig.20), Mandala strained and quirked, and other several Mandalas coming and going.

The audio part uses fragmented singing voices as its materials. The surface configurations of voices are the most

important, letting the significances and nuances of the voices come through. Depending on the configurations, the musical materials are first sorted into short sounds and continuous sounds. Second, the short sounds are categorized into five categories based on their intonations: high-low, low-high, high, medium, and low. After that, they are more finely sorted, depending on the intensity of the sounds. The musical texture of these materials is composed according to the appearance of Mandala in the visual part, meaning that the visual part leads the audio part in the production process.

Samsara: The audiovisual poem as a story

Samsara, a nine-minute audiovisual poem produced in November 2008, symbolically features the interpretation of the world of the Indian epic *Ramayana*. Unlike other audiovisual poems, it has a story. However, the story is not always transmitted to the audience because the symbolization originates from subjective views. Rather, I used the story in order to develop my own art inspiration.

I have been devoted to *Ramayana* and its performing arts for several years, especially the Cambodian shadow play *Sbaekthom*. *Sbaekthom*, coming up in the large outdoor screen by lighting with the palm shells, is very translunar (Fig.21). The traditional music for this shadow play seems to give the sense of color, and the atmosphere of the music at a slow tempo seems to give the sense of endless time. The music and the puppets of *Sbaekthom* are the main materials of this piece, and the landscapes of the Angkor ruins and area are just supplementary material (Fig.22).

The abstract of the *Ramayana* story in this poem is as follows; The Crown Prince of Kosala Rama, who was forced out of the country by his wicked stepmother, lived in seclusion with his wife, Princess Sita. The beautiful Princess Sita was kidnapped by the Ogre king Ravana and incarcerated on Ranka Island. Prince Rama declared war on the Ogre king Ravana to save his wife, Sita. With the help of the monkey general Hanuman, Rama went to Ranka Island and destroyed Ravana, and finally saved his wife. Some of Rama's retainers suspected Sita had not been chaste during her incarceration, and she jumped into the fire in order to prove her innocence.

Understanding this story is not necessary to enjoy this audiovisual poem. On the other hand, paying attention to the

relationship between visual and audio expressions based on the story gives the audience some more meaning to enjoy the poem.

Impatience: The audiovisual poem as moving fine art (2)

Impatience, an 11-minute audiovisual poem produced in November 2008, is composed of only one photograph. The photograph is one I took myself, of the relief of the Female God Devada, at Angkor Wat. Countless of statues Devada are found, not only at Angkor Wat, but also at other temples. No Devada is the same. Each of them has a very individual face and expression. This piece expresses Prince Rama's frustration as he fights the battle with the Ogre king Ravana to save his wife, Sita. His passion for Sita brings Rama to call Sita to his mind.

As for the composition, it is more like a Devada variation based on one photograph, using animation techniques to break up the photo into pieces or add image processing (Fig.23, Fig.24). The audio part only uses the C major scale and very limited additional material, like a cluster of segmented human voices, which can be varied by exchanging the materials or adding new concrete sounds.

The aim of this 11-minute using only one photograph is to prove that the visual part has the aspect of moving fine art.

This research was supported by a Grant-in-Aid for Challenging Exploratory Research (Research Project No. 22652021 “Structured approach to semantics and syntax of sound in movie”) from the Japan Society for the Promotion of Science (Nihon Gakujutsu Shinkoukai)



Fig.1: *Walk*



Fig.2: *Walk*

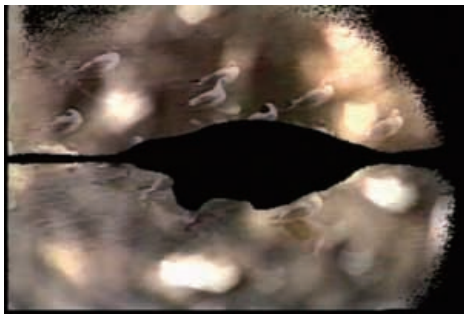


Fig.3: *Yin-Yang*



Fig.4: *Yin-Yang*



Fig.5: *Epitaph*



Fig.6: *Epitaph*

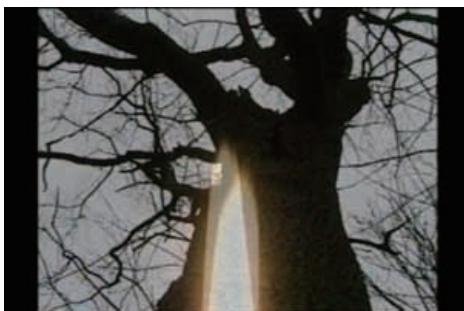


Fig.7: *sabi*



Fig.8: *sabi*

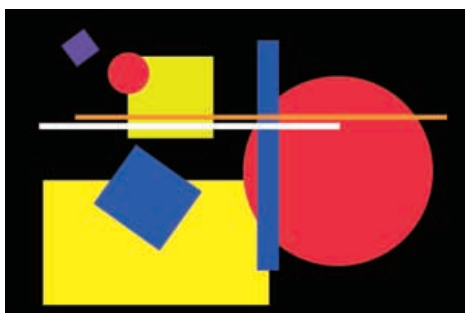


Fig.9: *Life*

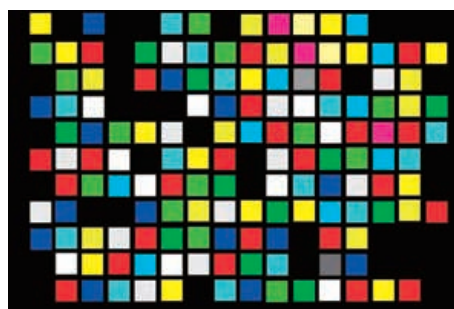


Fig.10: *Life*

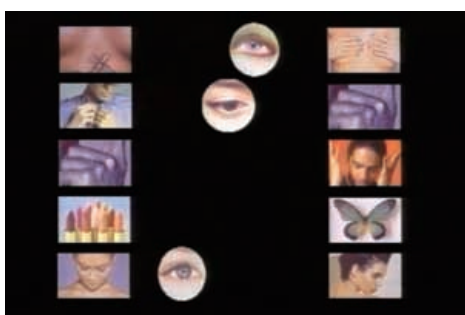


Fig.11: *Play*

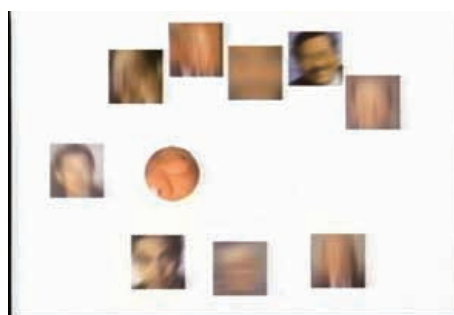


Fig.12: *Play*



Fig.13: *Common Tragedies in Urban Life*

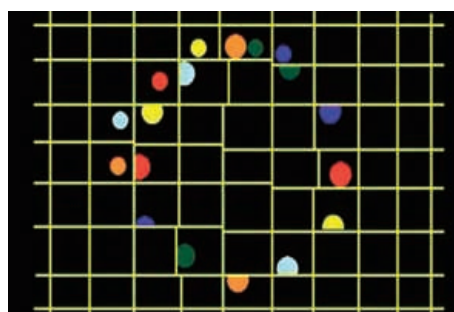


Fig.14: *Common Tragedies in Urban Life*



Fig.15: *Lust*



Fig.16: *Lust*



Fig.17: *Naga Variation*



Fig.18: *Naga Variation*



Fig.19: *Mandala Fantasy*



Fig.20: *Mandala Fantasy*



Fig.21: *Samsara*



Fig.22: *Samsara*



Fig.23: *Impatience*

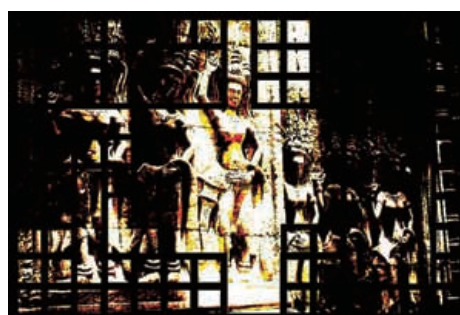


Fig.24: *Impatience*