SETTINGS AND THE SETTINGS STRUCTURE:
The Description and Automated Propagation of Networks for Perusing Videodisc Image States.

Alan P. Parkes

Science and Engineering Research Council

Information Technology Research Fellow

Dept. of Computing

University of Lancaster

U.K.

ABSTRACT.

This paper describes a system for formally representing spatial relationships between videodisc image states called *settings*. A number of *setting relations* are defined, these being based on the manipulations of the camera typically used in the production of the moving film: *zooming in or out, panning* etc.. An algorithm is presented which, given a limited level of initial specification by a describer, will constrain, where possible, the setting relations holding between all pairs of settings. The resulting network is called the *settings structure*. The paper begins by placing the *settings structure* into the context of its being one part of the *CLORIS* system.

1. INTRODUCTION.

This paper describes developments in one area of an on-going project researching into an Artificial Intelligence (AI) approach to the *conceptual description, retrieval* and *discussion* of Computer Controlled Videodisc (CCV) material. CCV allows a *videodisc* (essentially a collection of some 50,000 individually addressable *still frames*) to be controlled by a computer system. Any frame can be accessed and displayed, and groups of frames can be displayed as *moving films*. Often, generated *graphics* and *text* can be overlayed onto the *video image*.

This paper describes an approach to the representation and automated propagation of *spatially oriented relationships* between *image states* called *settings*. The *relationships* are based on the typical *camera manipulations* of the moving film (*zooms, pans* etc.), and the resulting structure - called the *settings structure* - forms the basis of a system's intelligent manipulation, or a user's guided perusal, of the images from a *visual database*. The paper begins by providing a brief description of the wider context into which the work presented here should be placed. The (moving film and still image) construct the *setting* is then defined, and, with the aid of a simple abstract device, the *setting relations* are presented. The setting relations are then used to define the *settings* *structure*, along with an algorithm which, given a minimal level of specification by some *describer*, constrains, where possible, the number of setting relations which are applicable between all pairs of settings. An actual moving film is used as the basis of the examples.

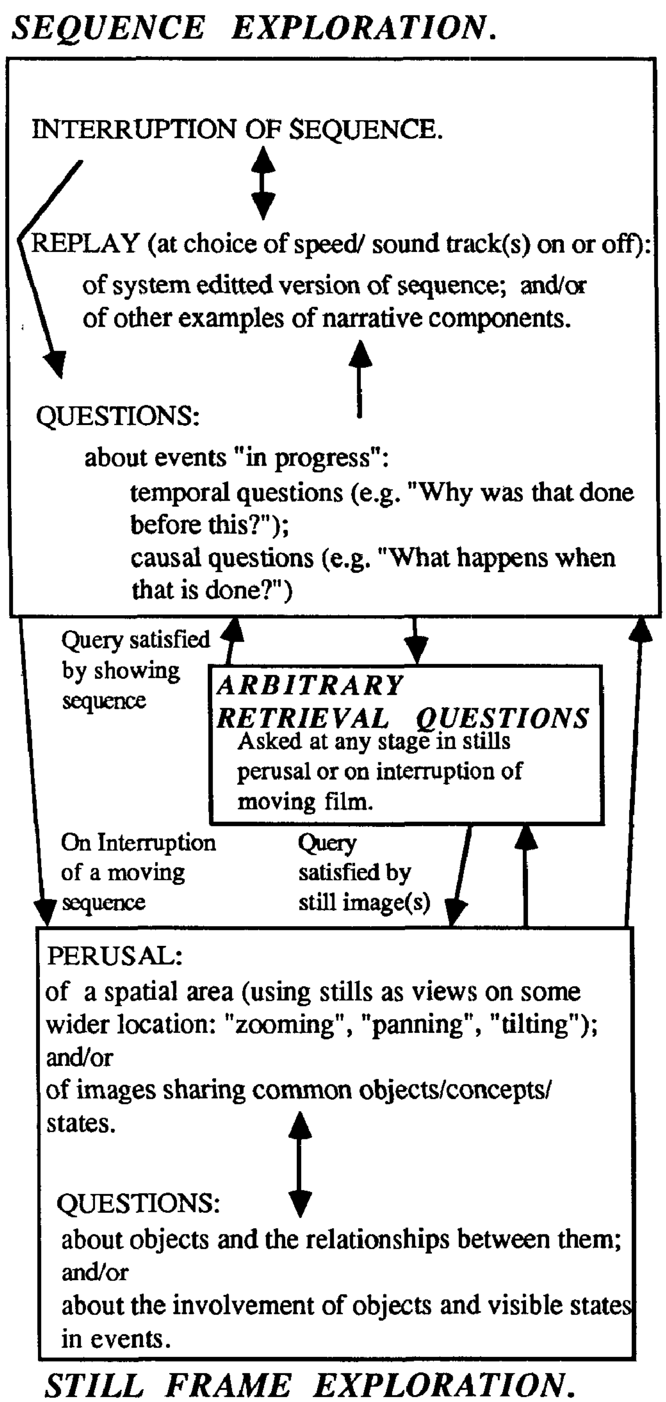


FIGURE 1. Computer Controlled Video as an Exploration Enabling Technology.

2. "CLORIS" - TOWARDS THE AI-BASED CCV SYSTEM.

CCV can be regarded as an *exploration enabling technology* (Parkes, 1988b). This means that the *user* will have control both over the display of the material, and over when and which questions are asked about it (see figure 1). As can be seen from figure 1, the scenario encorporates (a) *arbitrary retrieval questions* at any stage in the viewing; (b) *interruption of moving film sequences followed by questions about the events in progress and the objects on view*; and (c) *perusal and subsequent questioning about still frames* - whether or not the still frames are part of the moving film. The subject of this paper is the "perusal of a spatial area" component. However, a prototype system, which has been developed by the author (see Parkes, 1988a, 1988b), includes an implementation of these *spatial perusal* facilities. The prototype is called CLORIS (Conceptual Language Oriented to the Representation of Instructional film Sequences), and, to put the remainder of the paper into its proper context, this system will be briefly described here.

2.1. THE "CLORIS" ARCHITECTURE.

CLORIS provides an environment in which visual material (moving films and still pictures) can be initially *described*, and can subsequently be the subject of viewer *exploration* and *questioning* in ways described in figure 1. Figure 2 shows the *architecture* of the CLORIS system.

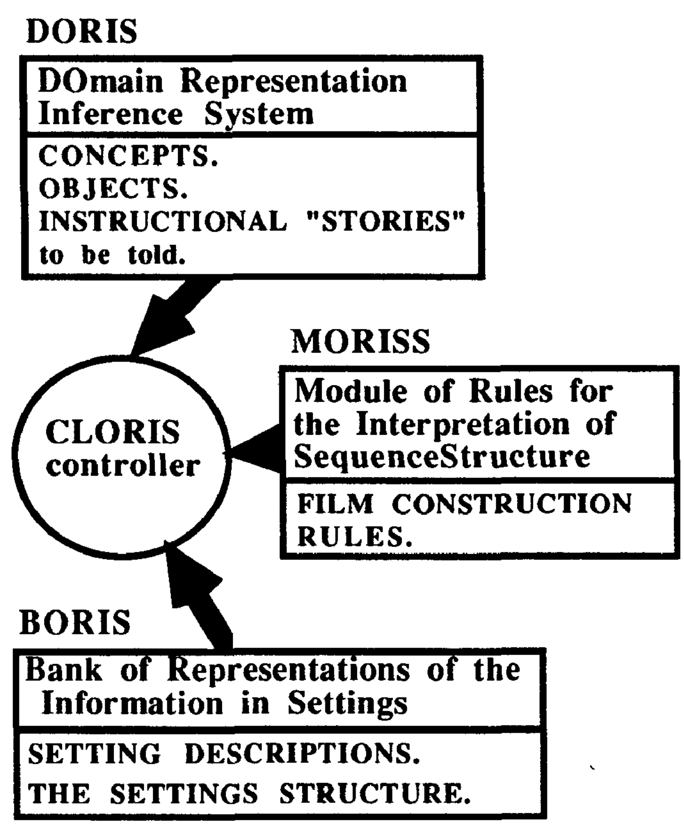


FIGURE 2. CLORIS System Architecture.

There are three *representation* modules, and a *controller* module which uses these (Parkes, 1988a and 1988b give more details about the modules themselves).

The DORIS module contains *abstract (i.e. "media-independent") descriptions of the relevant concepts, objects and events.* The constructs used to represent the information are based on a "hybrid" formalism derived from the "scripts" of Schank *et al.* (Schank and Abelson, 1977; Schank and Reisbeck, 1981), and "conceptual graph" formalism (Sowa, 1984). DORIS scripts (called "script ABSTRACTIONS") describe the *sequence of events which the performance of some task* (for example) *typically involves.* In a script abstraction, the *actors, props* and *locations* remain *uninstantiated,* these things being subsequently provided in connection with *specific realisations of the narratives* (see below).

The MORRIS module (at present undeveloped) is intended to contain *procedures for the production of multimedia "explanations"* i.e. using text, graphics, pictures, films and sound in coordinated and user-responsive ways.

The module called BORIS holds descriptions of the *particular images held in storage.* It is here that the *specific objects* defined in DORIS are related to *actual images.* BORIS contains both descriptions of, and structures detailing the spatial relationships between, the *described images* available. The former are in the form of ***setting descriptions***, while the latter is known as the ***settings structure*** - these are the main subjects of this paper, but first, the controller module of CLORIS is briefly described.

The CONTROLLER module accesses the three representation modules according to the user's current activity. The module also has access to "script INSTANTIATIONS" (also called "script APPLICATIONS") which show the mapping between the narrative specified by an abstraction and the *realisation of that narrative in film form* i.e. the *actor, prop* and *location* "slots" in the abstraction are *specialised* in the *application,* and the whole event sequence is related to a *specific piece of film which the describer has asserted to realise the narrative in question.* The events are mapped, in a one-to-many fashion, onto components of the *shot* (the portion of the film from "cut" to "cut"), called *settings,* which will now be described.

2.2. THE "SETTING".

The *setting* is essentially defined as being a group of one or more *still images* which display the same *objectively visible* "scene". Effectively, the *objectively visible* dimension of an image consists of the *physical objects* represented in an image and the relationships between those physical objects which can be *seen,* and are not *inferred-to-be-present,* as it were. Still images have a *continuum of meanings* (Parkes, 1988b) ranging from the *objectively visible* (what they objectively "show"), through *event-ambiguous,* where a member of a *set* of *events* could be assigned as being what the image could be inferred to be displaying a "moment" of, and finally to *event-determined,* where some other information beyond that contained in the actual image itself leads the viewer to infer that a *particular event* is represented (in "snapshot" form). The author's contention is that *descriptions actually applied to the image itself should be as objective as possible,* because it is only by doing this that the maximum flexibility of use of the arbitrary image can be guaranteed.

setting **F**
frames **3000 - 3096**
screen

logical

[THUMB:t1]. => <(0,0),(0,7),(23,14),
(25,13),(25,12),
(24,7),(17,0)>

[COLLAR:c1]. => <(6,17),(11,20),
(16,23),(28,13),
(22,6)> *priority* 1

physical

srels **(MZi)->[Fv].**

spatial

FIGURE 3. Example Setting Definition for Setting "F".

2.3. SETTING DESCRIPTIONS.

A setting, then, is an objectively-visible situation in which nothing much changes. Associated with the setting is the *setting description* (see figure 3), which has any combination of up to three dimensions:

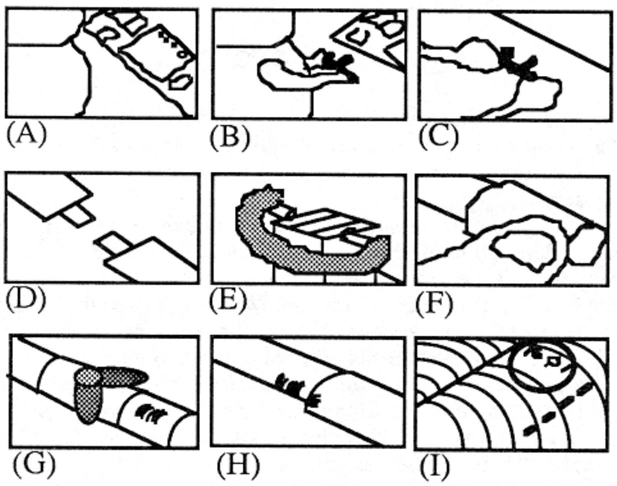
(1) *A propositonal account of what the image shows.* This contains descriptions of the *specific* objects in the setting and any *objectively visible* relationships between those objects. This might be called the *logical dimension.* When taken in conjunction with the (script) event descriptions, these propositions are used by CLORIS to derive statements describing what, about the event, can be *seen* at a given point in the moving film. Figure 3 describes a (simplified) setting taken from the *micrometer film* (see figure 4), when the engineer is tightening the collar - inference (when the film is interrupted in this setting): "the engineer is tightening the collar using his forefinger and thumb [*event*], and you can see the collar being tightened by the thumb" [*inference based on what is known to be visible*]. In itself, the setting is a view of a configuration of objects (e.g. the example setting shows the collar of a micrometer), and can be used as such, whether it is from a moving film or not. As has been discussed elsewhere (Gecsei, 1987), conventional keyword-based retrieval approaches are not suitable for image description and retrieval, particularly, as Gecsei points out, because information concerning the relations between the objects in the images is lost. This also leads Gecsei to to the use of conceptual graphs for image description. Moreover, the conventional approach also makes it very difficult to maintain the critical distinction between individual and type - a distinction that seems to have been omitted even by those researchers who have rejected this approach (Clark and Sandford, 1986; Gecsei, 1987; Halin *et al*, 1988; Lelu, 1988; McAleese, 1985). It is not sufficient for a system to know only that two separate photographs depict objects of the same type: it should also be able to determine whether they depict the *same actual object.* Thus, the object descriptors appearing in setting

descriptions are *individualised* i.e. denote *particular individual objects.*

(2) In what could be called the *physical dimension,* a setting description can also include *descriptions relating the objects in the setting to actual screen locations.* In CLORIS, these descriptions are in the form of *polygons* (initially drawn round the on-screen objects by the describer) - see Figure 3. Since a setting is a visual *state,* and not necessarily a *single frame,* (limited) movement of an object within a setting can be specified by *transformation* (i.e. scaling, translating and rotation) operations on the polygons.

(3) The third dimension of the setting description consists of *specifications of a subset of the setting relations holding between the setting and others.* This information is used by a propagation algorithm to produce the *settings structure,* which represents the available settings as constituting "views" on an overall (changing) visual situation. This dimension we could call the *spatial,* or *geographical,* dimension.

The last of the above three items is the main concern of the remainder of this paper.



The Micrometer Film:
The sequence opens with an engineer sitting at a work bench [A]. From a small case on the bench (screen right) he takes a micrometer, and after replacing the case he picks up a small cloth from the bench (screen left) and cleans the micrometer [B-C]. He then moves a small metal workpiece from the rear of the bench to the front of the worksurface (screen centre). He uses the micometer to measure the width of the workpiece [D-G] and writes the reading ("16.01") in a small exercise book (screen top left in [A]) [I].
NOTE that there are actually 39 settings in the micrometer film. The above drawings summarise the overall situation.

FIGURE 4. "Micrometer Film" Settings.

3. FORMAL RELATIONSHIPS BETWEEN WHOLE IMAGE STATES: THE SETTINGS STRUCTURE.

Images can be related because they depict the same object (i.e. the same object *constant* appears in all their descriptions), or objects of the same type (i.e. the same type label property appears in all their descriptions, or, more generally, a common supertype is shared by a type label in each description). Alternatively, the images could also depict *states* of the same type (or subtypes of a common supertype); likewise for events. There is, however, a further level at which relationships between images can be maintained, and this is at the level of *whole images.* Whether or not a still frame is taken from a database of still pictures, or happens to be taken from some (undisplayed) piece of moving film, or is an *interrupted* moving film, use can be made of facilities for the spatial perusal of single images *as if one were perusing some wider area upon which such images were "windows",* as it were. However, these "windows" would not simply allow perusal of an area in terms of moving up and down, or left and right, but also moving *in* to, or *out* from, a "scene". In the following sections, a formal basis is described for approximating the relationships between whole images in a way which will permit a structured (viewer's) perusal, with (system) assistance, of such images.

For CLORIS purposes, a film consists of *sequences* of *scenes,* where a *scene* is defined: *a succession of shots depicting a series of events taking place, in some location which holds constant, over a continuous period of time.* Such a *scene,* especially in the case of the video sequence, is often filmed using one or more fairly fixed camera locations (the micrometer film effectively uses only *one*). A *cut* will often involve either (a) transferring control to another camera, or (b) transferring instantaneously (by editing, say) to a *zoom* (in or out) of a previous setting. However, a stills database collection could also include *closeups* and *pans* of various stills. Particularly in the moving film, settings can be partitioned into structured sets, each set consisting of settings which are all *sub-settings* of some wider (possibly *virtual*) setting. The definition of the *scene* dictates that its settings will form such a structured set. Now, if a film is *filmic* in the sense meant by Carroll (1980) - an analogy of the linguistic concept "syntactically correct" - and at least one shot in a scene encorporates the entire geographical extent of the contents of the other shots in that scene, then clearly, all other settings will be sub settings of the geographical extent thus encorporated. The micrometer film is such an example (see figure 4) - all the actions take place in the area defined within the opening shot. We need make no demands on the strict correctness (if such a concept is applicable), of films, but it can be said that the more *filmic* the film is, the more fully will the facilities outlined henceforth be available.

3.1. THE BASIC SETTING RELATIONS.

It is helpful to our understanding to refer to the relationships between images in terms of the names given to the movement of the filming camera, or adjustment of its lens. Note that it is not the movement itself which is being referred to, but the *outcome* of that movement or adjustment - *how* the image was actually derived is of no interest. The intention is to view the relationships between settings formally. To these ends, a simple hypothetical device will be imagined, which will allow the perusal of images in such a way that the operations of the device can be used to define the relationships between settings. The device consists of a huge video screen, infinitely extendable in all directions, on which still photographs are displayed. Extension of the screen does not affect the resolution of the currently displayed photograph. One can view this screen through a fixed size (rectangular), window in a board which totally obscures all of the video screen except for the currently viewed portion. The board can be moved in any combination of perpendicular and horizontal directions, and the viewing window can be initially located anywhere on the screen. The *screen* can also be rotated about a point in the centre of the window, in accordance with the current location of the window.

The setting relations can now be defined (A and B are setting names):

Zoom In: *written* **A Zi B**, *spoken "A is a zoom in of B"*. The window remains where it is, **B** is the initial view. A is derived by expanding the screen by some non unit factor.

Zoom Out: *written* **A Zo B**, *spoken "A is a zoom out of B"*. (The inverse of **Zi**). The screen is contracted to produce A.

For the next six definitions, the size of the *video screen* is unaltered for any one definition.

Pan Left: *written* **A Pl B**, *spoken "A is a left pan of B"*. If B is the initial view, then A is obtained by a leftwards horizontal movement of the viewing window by a sufficient amount to remove **B** altogether from view.

Pan Right: *written* **A Pr B**, *spoken "A is a right pan of B"*.The inverse of **Pl**. A is produced by rightwards movement.

[Note: The above are *exclusive* pans - the *non-exclusive* ones will be discussed later. Similarly for *tilts*, below.]

Tilt Up: *written* **A Tu B**, *spoken "A is an upwards tilt of B"*. If B is the initial view, then A is obtained by an upwards perpendicular movement of the viewing window by a sufficient amount to remove B altogether from view.

Tilt Down: *written* **A Td B**, *spoken "A is an downwards tilt of B"*. The inverse of **Tu**. A is produced by *downwards* movement.

Roll Clockwise: *written* **A Rc B**, *spoken "A is a clockwise roll of B"*.If B is the initial view, then A is obtained by clockwise rotation of the window about its exact centre.

Roll Anticlockwise: *written* **A Ra B**, *spoken "A is an anticlockwise roll of B"*. The inverse of **Rc**. A is produced by anticlockwise rotation.

Note that *rolling* is virtually unknown in "CLORIS films" (it is rare in all films!). It is included in the interests of completeness.

3.2. A "WINDOW" ON A CHANGING WORLD: MODIFIED SETTING RELATIONS.

Since the setting relations are also meant to apply to settings taken from the moving film, there are other factors which need to be considered. Although some settings in the moving film are clearly (zoom, pan, tilt) related to one another, there may be minor differences between them caused by the dynamism of the events taking place in the film (this could also apply to the stills database). The visual state is very likely to have been *modified*, in some ways, but insufficiently to mean that the relationship between particular affected settings has been substantially altered. The device screen can be used to present a series of photographs representing the (ordered) moments from some event. Let us assume that one "snapshot" is overlayed perfectly onto the previous one, so that leaving the viewing window still will produce the same view of all static objects and locations, whichever snapshot is being viewed. This allows the following definition to be made:

Modification: *written* **A M B**, *spoken "A is a modification of B" or "B is a modification of A"*. **B** is the initial view, and A is derived by (a) perfectly overlaying a photograph over the one which **B** views so that all entities unmoved and unchanged are in exactly the same locations as they were initially, and (b) some of the changes to the entities are contained in A. Note that condition (a) alone, and where *no* objects move, defines the equality relationship (=) for settings.

It is clear that modifications can take place which affect the nature of zooms (by definition, non-exclusive), and *non-exclusive* pans and tilts. To cater for such phenomena, the relation **M** can be *composed* with any of the basic relations, such that, if **R** is one of these, then **MR** is the name of the form of **R** which *includes* modification. In these cases, sufficient of the initial view remains in the new one so that the modifications can be *seen* - if it did not, then the **M** could simply be dropped. (A **MR** B is spoken *"A is a modified 'R' of B"* - thus A **MZi** B = *"A is a modified zoom in of B"*).

The situation involving non-exclusive panning and tilting is slightly more complex. However, panning and tilting are *isomorphic* activities, so it is necessary to discuss the panning group only. Moreover, since **Pl** and **Pr** are also clearly isomorphic activities, **Pl** alone will be discussed. The definition of **Pl** omits the possibility of an *non-exclusive* pan: when the viewing window is moved insufficiently to remove *all* of the initial view from the window. First, the non-exclusive pan (without modifications) is considered. This can be defined as follows:

Overlap Pan Left: *written* **A OPl B**, *spoken "A is an overlapping (or non exclusive) pan left of B"*. If **B** is the initial view, A results from moving the window leftwards by a distance which is less than the width of the window.

The modified form is defined:

Modified Pan Left: *written* **A MPl B**, *spoken "A is a modified pan left of B"*. As one would expect, A is derived as is B in the above description, but *first* **B** is *modified* and A contains a part of **B** in which some of these modifications can be seen to have been made.

Of course, several setting relation can apply simultaneously between settings (a zoom in pan left, for example). Here, we are interested first in describing formal properties of the setting relations individually. The more complex situations will be discussed later.

(a) ZOOM TABLE.

[If X, Y, and Z are settings, and X r1 Y, Y r2 Z, where r1 and r2 are setting relations, looking up table entry with row labelled r1, column labelled r2 will give the setting relations holding between X and Z - this also applies to (b) below.]

a \ b	M	Zi	Zo	MZi	MZo
M	M, =	MZi	MZo, Zo	MZi, Zi	MZo, Zo
Zi	MZi, Zi	Zi	Zo, =,Zi	MZi, Zi	all
Zo	MZo	Zi, =,Zo	Zo	Mzo, M,MZi	MZo
MZi	Mzi, M	Mzi	all	MZi, Zi	all+
MZo	MZo, Zo	MZo, M,MZi	MZo, Zo	all+	MZo, Zo

NOTE:
"all" = {M,Zi,Zo,MZi,MZo};
"all+" = all U {=}.

(b) PAN (and TILT) TABLE.

a \ b	M	Pl	Pr	OPl	OPr	MPl	MPr
M	M, =	Pl	Pr	OPl, MPl	OPr, MPr	OPl, MPl	OPr, MPr
Pl	Pl	Pl	all+	Pl	OPl, Pl,MPl	Pl	OPl, Pl,MPl
Pr	Pr	all+	Pr	OPr, Pr,MPr	Pr	OPr, Pr,MPr	Pr
OPl	OPl, MPl	Pl	OPr, Pr,MPr	OPl, Pl	=,OPl, OPr	OPl, MPl,Pl	M,MPr, MPl, OPl
OPr	OPr, MPr	OPl, Pl,MPl	Pr	=,OPl, OPr	OPr, Pr	M,MPl, MPr, OPr	OPr, MPr,Pr
MPl	OPl, MPl	Pl	OPr, Pr	OPl, MPl,Pl	M,MPl, MPr, OPr	OPl, MPl,Pl	=,M,OPr, MPr,OPl, MPl
MPr	OPr, MPr	OPl, Pl	Pr	M,MPr, MPl, OPl	OPr, MPr,Pr	=,M,OPr, MPr,OPl, MPl	OPr, MPr,Pr

NOTE:
For table for TILT relations, substitute the string "Tu" for the string "Pl", and the string "Td" for the string "Pr" (or vice versa) THROUGHOUT the above table.
all+ = {M,=,Pl,Pr,OPl,OPr,MPl,MPr}

FIGURE 5. Setting Relations Transitivity Tables.

3.3. THE DEFINITION OF THE SETTINGS STRUCTURE.

A collection of settings and the above setting relations can now be defined as a structure.

Definition: *A structure SS = <S, R>, where:*

S is a non-empty set of settings;
R is the set of setting relations;

for each $S_a \in S$, *if* $S - \{S_a\} \neq \{\}$
then there exists
$S_b \in S - \{S_a\}$ *such that for some* $n \geq 1$:
$S_a\ \underline{R_1}\ S_1,\ S_1\ \underline{R_2}\ S_2, .., S_{n-1}\ \underline{R_n}\ S_n$
where $S_n = S_b$ *and* $S_i \in S - \{S_a\}$
$R_i \in R,\ 1 \leq i \leq n$;

is called a ***settings structure.***

[i.e. either any two settings are directly linked, or there is a path from one to the other through other settings].

A settings structure can be drawn with the nodes as setting names (or schematic drawings of the settings), and the directed arcs labelled with the appropriate setting relation. See figure 8 for part of the settings structure of the micrometer film.

3.4. DESCRIPTION AND PROPAGATION OF SETTING RELATIONS.

The description of a setting can include a specification of which other setting(s) that setting is related to. It would, however, be an extremely time consuming activity to specify all possible relationships between any two settings in this way, and, fortunately, this is unecessary. Take *zoom in*, for example. This relation has certain properties (i.e. transitivity, irreflexivity, and antisymmetricality), which also applies to *zoom out* and the basic *panning* and *tilting* relations. In these cases, therefore, the responsibilty of the describer would be to supply sufficient specifications to ensure that the propagation of all the relations can be carried out automatically. The relation **M** is *not*, in general, transitive (though it is clearly symmetrical and irreflexive) unless an assumption about inequality is made: if A **M** B and B **M** C for settings A,B, and C, if the assumption is made that no two settings are equal unless explicitly marked as such, then A **M** C is a valid inference. This assumption is necessary because the changes done to A to produce B may have been *undone* to produce C from B, thus resulting in C being identical to A. This *in*equality assumption also applies to the relations which use **M** in their definitions i.e. **MZi** etc.. If A **MZi** B and B **MZi** C, then, in general, it is not strictly true to state A **MZi** C *alone* because the **M** part may have been *reversed*, which would mean that A **Zi** C. It could be an assumption of the system, though, that unless, explicitly stated (by the describer) that A **Zi** C, the single **MZi** relation holds. Then, the effects of modifications could be assumed to be unreversed, unless explicit indications to the contrary exist. Figure 5(a) shows the transitivity table for the *zoom* relations.

```
addrel R(i,j) =
add <i,j> to queue ToDo;
while ToDo not empty do
begin
 get next <i,j> from ToDo;
 N(i,j) := R(i,j);
 for each node k such that
    compatible(N(k,i),R(i,j))
  begin
     x := N(k,j) ∩ constraints(N(k,i),R(i,j));
     if x ⊂ R(k,j)
     then R(k,j) := x
     endif;
     if x ⊂ N(k,j)
     then add <k,j> to ToDo
     endif;
  end;
  for each node k such that
      compatible(R(i,j),N(j,k))
    begin
      x := N(i,k) ∩ constraints(R(i,j),N(j,k));
      if x ⊂ R(i,k)
      then R(i,k) := x
      endif;
      if x ⊂ N(i,k)
      then add <i,k> to ToDo
      endif;
    end;
end.
```

```
constraints(r1,r2) =
C := {};
for each x in r1 do
   for each y in r2 do
      C ∪ T(x,y)
   end
 end;
 return (C)
```

Notes:

(i) N is the existing network, such that N(x,y) is a SET of setting relations holding between the settings x and y. Note that N(x,y) is the set of ALL setting relations if no entry exists.

(ii) R is a copy of N with ALL elements set to the set of ALL setting relations EXCEPT for R(i,j), which holds the new relation between settings i and j to be added to N.

(iii) On exit, N holds the updated network.

(iv) The table T in "constraints" is set to be the appropriate transitivity table i.e. "ZOOM" , "PAN" or "TILT".

(v) "compatible" - ANY setting relations are compatible with "=" or "M". Otherwise two sets of relations r1 and r2 are compatible iff ALL members of r1 - {=, M} and ALL members of r2 - {=, M} are in the same (i.e. "ZOOM" or "PAN" or "TILT") GROUP.

FIGURE 6. The Propagation Algorithm.

Such a table can be used to fill in the complete *zoom related* network for a particular set of settings, given a minimal level of specification by a describer - as will be discussed shortly. Figure 5(b) shows the transitivity table for the *panning* (and *tilting*) groups.

The above setting relations are not the only ones that could have been chosen. *Pans* (and *tilts*) for example, could be represented so that the system "knows" that one setting "abuts" another, so to speak. The point is that, if the setting relations are selected appropriately, the same algorithm (figure 6) will handle the propagation of the relations throughout the network when a new relationship between two settings is added to a settings structure. Figure 7 shows a simple example of how a particular settings structure is updated by the algorithm. In fact, the algorithm was derived from Allen's (1983) *temporal relations propagation algorithm*, the only changes being made because, due to what must have been typographical errors, *his* algorithm did not operate correctly!

3.5. KNOWING WHERE YOU ARE - VIRTUAL SETTINGS.

As previously stated, a property of (film) scenes which obey Carroll's *included shot space constraint*, in which the whole spatial bounds of a *scene* are encompassed by (at least) one *shot*, is of particular importance for the nature of the settings structures of those scenes. In fact, if the property is restricted by substituting *setting* for *shot*, in the preceding sentence, then it better suits the purposes here. Suppose A is such an *encompassing* setting. Any other setting, S, from the scene will be such that A **M** S, A **≡** S, A **Zo** S, or A **MZo** S. In other words, in all cases except **M** or **≡**, all the settings in the scene are *zoom ins* of A. Such a structure has useful implications for system-assisted user browsing of related images.

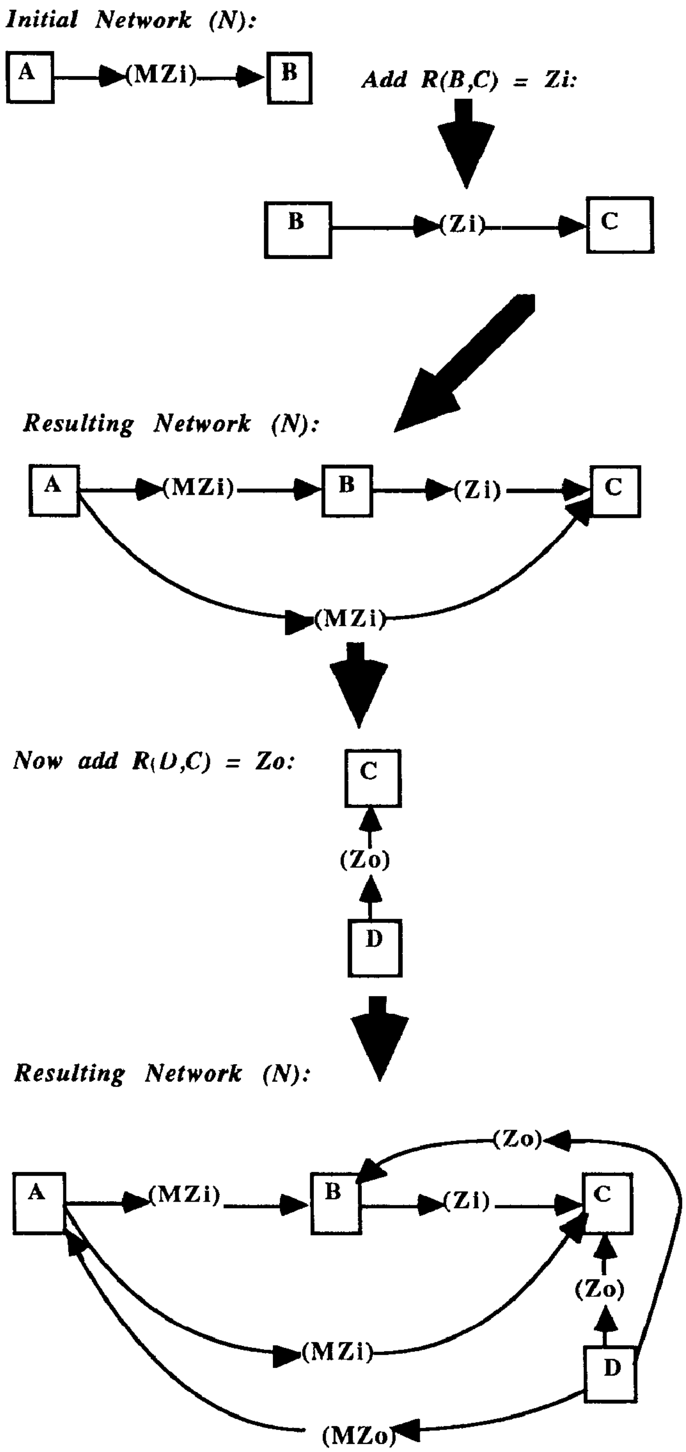
The abstract device is, of course, an *approximation* of the real situation. In the transition between two settings, more than one relation could apply between those settings. In the moving film, this occurs when, for example, the camera pans left *at the same time* as it is zooming in. As things stand, the settings relations will not suitably handle these cases: the algorithm will update the structure with respect to *subnetworks* in each of which all settings are *zoom related*, or *pan related* or *tilt related*. This is one reason why the *encompassing setting*, hereafter called the *main* setting, is such an important concept. However, the situation is not quite as simple as was indicated by the previous paragraph. All settings (which are not equal to or modifications of the main setting) are not likely to be strictly zoom-in related to the main one. To illustrate this, consider the original settings structure for the micrometer film, on which figure 8 is based. This had the following entries:

I **MZi** B and C **Zi** B.

(with the entries I **Zi** A_v, A_v **OTu** A_m not present). Using the table in Figure 5(a) to constrain the relationship between I and C we look up the entry for **MZi**, **Zo** - the latter being the *inverse* of **Zi** - which yields the following possibilities:

I **M** C or I **MZo** C or I **MZi** C or I **Zo** C or I **Zi** C .

From Figure 8, it can be seen that none of these is in fact the case. I depicts a 'scene' which is (in A) *above* the part of A which is occupied by C.



Notes:

(i) Taking a constraints path into arrow heads involves inverting the setting relation labels encountered.

(ii) [B]->(Zi)->[C] = R(B,C) = Zi = "B is a Zoom In of C"

FIGURE 7. Operation of the Constraints Algorithm.

By propagation, this situation would "infect" the inferred relationships between **I** and **D**, **I** and **E** etc.. What is required is that these erroneous inferences are prevented i.e. that a setting **I'** is related to **I**, and to **B**, as follows:

I Zi I' and I' OTu B.

That is to say that **I** is really a *zoom in* of an image which is an *overlapping tilt up* of **B**. This "shields" **I** from **C** and thus inhibits the propagations which will be made regarding their relationships. Unfortunately though, **I'** does not exist (because the *zooming* and *tilting* took place at the same time). The solution to this problem is to allow **I'** to be entered (by the describer) as a setting applying to no actual images at all, but as nevertheless having a location in the *settings structure* - it is a ***virtual setting***. There is an additional problem, however: if **I' OTu B** and **B MZi A**, then, as things stand, the system does not know whether or not **I** is actually displaying something (in zoom in form) which is not even displayed in **B**. The solution to this is to *force* the system to *infer* that **I** *is* displaying a part of **B** by designating **B** as a *main* setting and introducing a rule which says: *any setting derived from a main setting directly by overlapping pans and/or tilts necessarily depicts spatial areas entirely encompassed by that main setting.* Thus **I** becomes a *zoom in* of **B** in this less restricted sense i.e. a *zoom in of the upper portion of* ***B***. Indeed, all settings which are linked to the settings which are directly *overlapping pan or tilt related* to **B** constitute a network representing a *partial perusal* of the spatial extent of **B**, in which no setting depicts any area outside of **B**. **B** is, of course, a (modified) zoom in of **A**, so the settings which are part of **B** are part of **A**, and any setting *directly zoom-in related* to **B** takes its *normal* place in the remainder of the settings structure i.e. is subject to any constraints which the algorithm imposes.

By following the above guidelines, the settings structure is effectively sub-divided into networks, each of which contains settings depicting one part of a main setting. To illustrate this, figure 8 shows the refined settings structure for the micrometer film. To make things simpler, **B** is not designated as a *main* setting, but, in fact, **I** has been directly linked to $\mathbf{A_v}$, which is a *virtual setting* representing an *upwards overlapping tilt* of **A**. Since **A** is a main setting, **I** necessarily only depicts an area entirely encompassed by **A**. The network with nodes **I**, $\mathbf{I_1}$, and $\mathbf{I_2}$ is constrained as it would be by the algorithm, but does not result in constraints being placed on the other parts of the settings structure. Setting names are subscripted with **v** if they are *virtual*, **m** if they are *main*. Note that the setting $\mathbf{C_{vm}}$ is a *virtual main* setting. All the real settings linked via the overlapping relations to this one constitute views of the micrometer. In perusing such a network, the system can keep track of the position of a particular view relative to its (immediate) main setting: $\mathbf{E_v}$ designates the left-side views of $\mathbf{C_{vm}}$, which are exclusive of parts of those designated by $\mathbf{F_v}$, whereas $\mathbf{H_v}$ and $\mathbf{G_v}$ designate right-side views which are possibly inclusive of some aspects shown by $\mathbf{E_v}$ views (the algorithm gives the relations **≡**, **Opr**, and **OPl** as the possible relationship between $\mathbf{F_v}$ and $\mathbf{G_v}$ - the describer could always check the results of the algorithm and identify which single relationships held, as appropriate).

Since, as stated before, it is desirable for there to be a *main setting* for the ***whole*** settings structure in question, if there is no such setting, then a *virtual setting* can be designated as the main one. A further point is that virtual settings, just as real settings, could have associated *setting descriptions* (minus the *physical* *dimension*). This could be useful if an inheritance system was used for descriptions i.e. one describes the *main* setting in great detail, and the system applies the appropriate parts of *that* description to the *zoom ins* of the main setting, by, perhaps, filling in some skeletal description, of such *zoom in related settings*, provided by the describer.

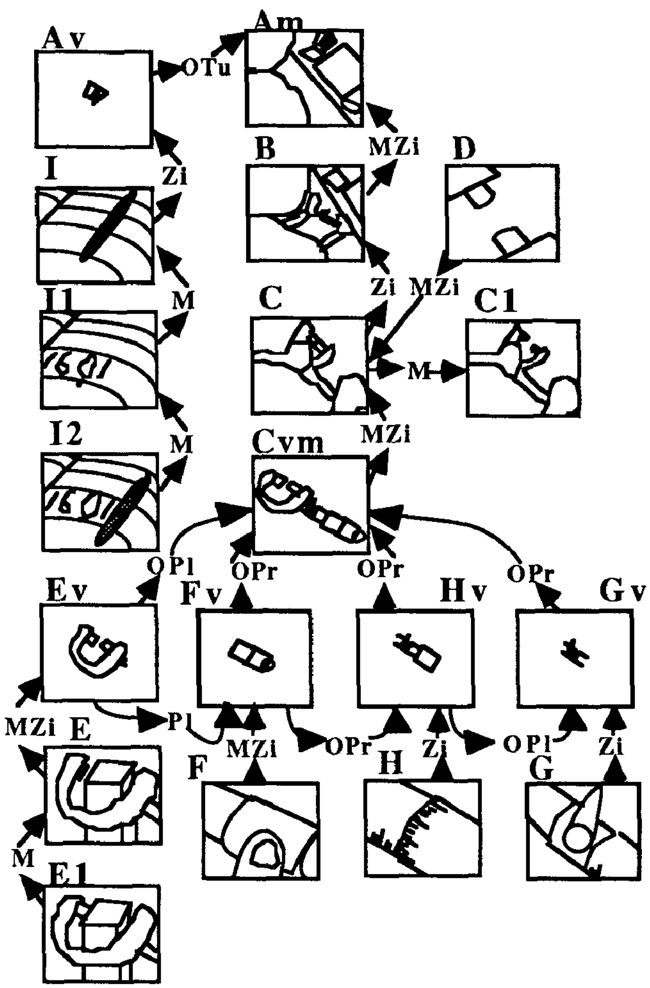


FIGURE 8. Refined Version of Micrometer Film Settings Structure

4. THE PURPOSE OF THE SETTINGS STRUCTURE.

The *settings structure* has several potential uses. The effects of *spatial deletion* (essentially the disappearance of objects from the screen, over time - Carroll, 1980) may be to remove an object from the screen before the user has had chance to see the location of that object. In the micrometer film, the book disappears from the screen in the display of setting **B**, only to reappear in setting **I**, when its location is not clear, since the book is shown in *extreme closeup* and occupies the entire screen. Since setting **I** is a zoom in of part of **A** (by the *main setting* rule), **A** should contain more information about the book's location than does **I**. In fact, in this case, since **A** is actually the *main setting* for the entire *scene*, and is not *virtual*, it shows the initial location of all the objects.

At any stage, if the currently viewed image (i.e. *frame*) is part of a setting which has a place in some *settings structure*, the selection of the next image to view can be made from sets of images of which each is related to the current one in some known spatial way. Moreover, the choices can be *ordered* so that the next image *zooms in* (for example) a little more than the previous one, but a little less than the following one etc.. These facilities, based simply on the designated relationships between whole images can be realised without recourse, by the system, to the actual *content* (i.e. *logical* descriptions) of the images. Furthermore, the describer is not involved in designating each possible path through the material, but rather all *potential* paths arise from the nature of the settings structure, and can thus be chosen *dynamically*, by the system or user, at any stage, according to the requirements of the interactive session. This level of browsing is available to the appropriate system, *whether or not the contents of the settings are logically described.*

A much more intelligent use can be made of a settings structure if the settings *are* described, however. If one setting is a modified version of another, it may be important to explain how that modification came about, or what it actually *is*. Similarly, it might be important to know what a zoom in is actually zooming in *on*. A person would be understandably nonplussed if he or she requested a zoom in on a "scene" displaying the micrometer, to see a close up of it, and was presented with a picture of the engineer's (empty) hands, because the modification was that the engineer had put the micrometer down somewhere! Nevertheless, the settings structure provides a way of reducing the number of target images in such situations.

The settings structure is only a partial solution to the problem of identifying the spatial relationships between whole images. For example, the panning and tilting discussed here is of a *two dimensional* nature. Sometimes a camera rotates about an imaginary perpendicular (in effect, this is what *panning* is: we have been discussing *tracking*). Similarly, the camera can actually *tilt* about an imaginary horizontal. In reality, then, the situation is highly complex, sometimes continuous, and many setting relations can hold, simultaneously, between two particular settings. Nevertheless, what has been attempted here is to provide a simple, *workable* basis on which images can be manipulated according to spatial considerations. The *setting description* handles the cases in which one requests an image related by some *conceptual relationship* to the present one. The settings structure is meant to handle the case in which one wishes to inspect an image which is *spatially* related to the present one in some specified way.

5. SUMMARY.

Images (i.e. *settings*) can be organised in terms of the spatial relationships between *whole* images (based on *pans, zooms, rolls*, and *tilts*) in such a way as to facilitate the construction (by system and/or user) of *dynamic* (*storage independent*) paths through the material. These *paths* can be used to peruse some entire area in an organised way (moving left, closing in, etc.). A set of settings can form a *settings structure* when at least one setting relation holds between any pair of settings in the set. Thus the description of a setting can include (or even *consist of*) a specification of which settings are related to it. The relational properties of the setting relations allow for a *propagation* algorithm to update a settings structure when a new relationship between two settings is added, thus rendering it unnecessary for the describer to enter all relationships between all settings at description time. The optimal form of a settings structure occurs when one setting (called a *main setting*) defines the entire spatial extent not transgressed by any other setting in the structure. Since the setting relations are approximations to the real situation (in which, say, zooming and panning can occur simultaneously), the idea of the *virtual setting* (serving as a link between "real" settings) is introduced. This prevents erroneous relations from being allocated by the algorithm in cases in which zooming is not performed *centrally* on one image to produce another. Moreover, if the system can assume that any directly overlapping pans and tilts on a *main setting* do not depict areas *outside* of those defined by that main setting, then any subsequent settings *zoom-in related* to those settings represent details of the *main setting*. Thus, a settings structure is partitioned into *subnetworks*, each of which can be used to peruse a designated area of the *main setting*, and from which the rest of the settings structure is protected (in the sense of the prevention of erroneous constraints being allocated by the algorithm). A system could make use of the settings structure without possessing any actual *setting descriptions* beyond those specifying the setting relations between settings, but optimal effectiveness would ensue from the use of *logically* described settings in a settings structure.

6. CONCLUSIONS.

The formal basis described in this paper can be implemented using fairly modest programming facilties, and it is a trivial matter to devise a "parser" for the entered relations between settings. CLORIS uses "pop-up menus", which can be invoked whenever a still image is being viewed These menus present only those perusal options available for the currently viewed image, in terms of selections such as "CLOSER VIEW" (*zoom ins*), "UP" (*tilt up*) etc.. Along with the *logical dimension* of the setting descriptions, and suitable *domain knowledge*, other views can also be classified with regard to the *objects represented in the image* and not merely the whole "scene" shown by it.

Part of the author's present research project involves investigating the suitability of the selected *setting relations*, and the feasibility of modelling more realistically the situations in which several setting relations apply simultaneously.

REFERENCES.

Allen J.F. (1983) Maintaining Knowledge about Temporal Intervals. in *Communications of the Association for Computing Machinery*. 26(11), 832-843

Carroll J.M. (1980) *Toward a Structural Psychology of Cinema*. Mouton, The Hague.

Clark D.R. and Sandford N. (1986) Semantic Descriptors and Maps of Meaning for Videodisc Images. *Programmed Learning and Educational Technology*. 23, 1.

Gecsei J. (1987) Browsing Techniques in Image Databases. Personal Communication. *I.R.O., University of Montreal*, Montreal, Canada.

Halin G., Mouaddib N., Foucat O., and Créhange, M. (1988) Semantics of User Interface for Image Retrieval. In conference: *Recherche d'Information Assiste par Ordinateur (RIAO '88), March 1988*, Boston, MA.

Lelu A. (1988) Browsing Through Image Databases via Data Analysis and Neural Networks. In conference: *Recherche d'Information Assiste par Ordinateur (RIAO '88), March 1988*, Boston, MA.

McAleese R. (1985) Knowledge and Information Mapping and Interactive Video. *Unpublished Booklet, University Teaching Centre, University of Aberdeen*, Aberdeen, Scotland.

Parkes A.P. (1988) The Prototype CLORIS System: describing, retrieving and discussing videodisc stills and sequences. to appear in *Information Processing and Management*. 24,5.

Parkes A.P. (1988) An Artificial Intelligence Approach to the Conceptual Description of Videodisc Images. (Unpublished Doctoral Thesis) *University of Lancaster, Lancaster*, England.

Schank R.C. and Abelson R.P. (1977) *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum, Hillsdale, NJ.

Schank R.C. and Reisbeck C. (1981) *Inside Computer Understanding*. Lawrence Erlbaum, Hillsdale, NJ.

Sowa J.F. (1984) *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, Reading, MA.

ACKNOWLEDGEMENTS.

This paper represents research done mainly while the author was engaged on his Ph. D. project, and was sponsored by the *Science and Engineering Research Council* and *Logica (Cambridge) Ltd.* The author is now sponsored by the *Science and Engineering Research Council.* Special thanks and love to *anna* and *betty.*