Behavior Sampling: A recording mechanism for visually based teleoperation.

Stephen PALM Taketoshi MORI Tomomasa SATO
Email: {palm, tmori, tomo}@lssl.rcast.u-tokyo.ac.jp
Sato Lab., RCAST, The University of Tokyo
4-6-1, Komaba, Meguro-ku, Tokyo 153-8904, JAPAN

Abstract

This paper proposes a visually based teleoperation with accumulation method and system where the visual and control information sequences are stored and available for syntactically indexed playback. Visually based teleoperation systems heretofore described in the literature unfortunately can only display an instantaneous representation of the control sequence. Past control experience should instead be accumulated and available to facilitate teleoperators. In this paper, the behavior sampling extraction theory and method, the behavior sampling data representation, and portions of the status on demand function-set are introduced. Behavior sampling utilizes semiotic analysis of the teleoperator control behavior to syntactically segment the control motions and the relevant visual objects into a syntactically indexable storage stream. A preliminary system used for manipulating single biological cells under an optical microscope (fitted with a video camera) is described. Repetitive manipulation experiments on Mato fluorescent granular perithelial (FGP) cells show the effectiveness of this enhancement to the visually based control approach.

1. Introduction

Visually based teleoperation control methods have been established for advanced master slave teleoperation. The visually based control methods 1) offer a more intuitive human machine interface and 2) allow for much simpler and economical control algorithms.

Visually based methods combined with teleoperation allow the goal condition setting (and some aspects of the path planning or constraints) to be set almost intuitively by the human while the automated visual feedback system executes the tedious control details. For example, the status driven microhandling system (SD-MHS) provides a robust means of manipulating sub-millimeter scale objects [1]. This visual communication interface (VCI) based teleoperation method is commensurate with the extensive work in the autonomous control area of visual servoing and control [2, 3, 4, 5].

As we look to other applications of visually based control, deficiencies in existing control systems become apparent. For example, semi-repetitive task execution such as biological cell handling where hundreds or thousands of cells must be processed individually, require the operator to repetitively setup the system.

Recent studies of aging and high fat diets have focused on analyzing Mato fluorescent granular perithelial (FGP) cells [6]. New techniques to analyze individual cells require the isolation of each cell by removing the tissue surrounding the cell. Figure 1 shows a visually based manipulator with a two micrometer wide scraper made of glass. The manipulator scrapes the undesired tissue from around the Mato FGP cell in preparation for its removal.

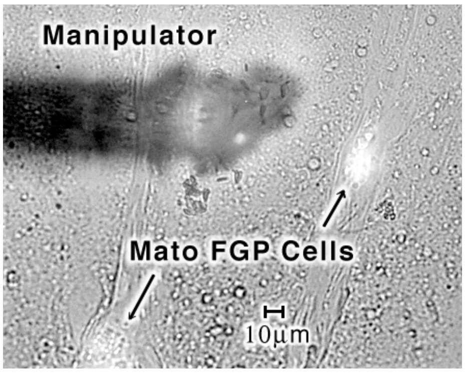


Figure 1. Cell manipulation environment.

Our experience with cell manipulation and other microworld tasks has highlighted several limitations in visually based control. These limitation can be grouped into three general areas.

1) The system can only display an instantaneous representation of the control sequence. The systems do not provide a means for displaying past control sequences. The only method of reviewing a past manipulation is to video tape the combined video image and graphical overlay of the control indications. Thus a data representation and storage mechanism is needed for visually based control mechanisms.

2) While some knowledge based or autonomous robots do have mechanisms for accumulating past experience, they tend to be based purely on image frame data or abstract representations or models. A means to accumulate the underlying raw data (both object visual representation and control instructions) is needed.

3) The SD-MHS system does not have an underlying concept or representation of the actual objects in the image being manipulated. Semantic information about the objects is only implicitly known by the human operator during sensing point selection. Likewise, in other visual servoing systems, a semantic model of the objects being manipulated may be known, but the motion functions are not classified syntactically or semantically. Thus, the ability to syntactically organize the information and allow

for the addition of semantic information should be improved.

As is evident from these limitations, an overall improvement of visually based teleoperation would come from the addition of a means to accumulate and display the raw visual and control data. A database of past motion sequences and their associated imagery would benefit both the human operator and the system. Humans are acutely communicative by visual means, thus visual representation of control sequences would be easy for the human to understand. For the system, accumulated knowledge in the form of raw segmented visual control sequences would allow reanalysis of the experiences for reuse in new control situations. Concepts from the field of hypermedia and video object-based compression offer some insight to a solution and will be reviewed and extended in Section 2.1.

Considering the above limitations, the *bilateral behavior media* cosmos between humans and their tools is born. The bilateral behavior media cosmos comprises three areas: 1) A data representation and extraction method for accumulation of visually based interactions between humans and tools. This is termed *behavior sampling* and will be discussed in detail in Sections 2 and 3. 2) A control methodology for visually specifying and visually controlling teleoperation/machines. The paradigm is referred to as *status driven* and was presented in a previous paper [1] and is summarized in Appendix I. 3) Functions for assisting and supporting humans through visual mechanisms. Capabilities include visually navigated "redo" or "undo" based upon past visual-control sequences. This functionality is expressed as *status on demand* and introduced in Appendix II. Together the three areas encompass the notion of bilateral expression of behavior between humans and machines through a multiplicity of visual media.

Three areas of behavior sampling will be developed in this paper as follows: 1) A structured <u>data representation</u> for visually based control information and the associated objects/images. The data representation is syntactically based allowing manipulation at various levels. The data representation will be theoretically discussed in Section 2.2 and the implementation will be discussed in Section 4.1. 2) <u>Methods</u> for extracting (behavior sampling) and interjecting those elements of visually based telerobotic control will be introduced in Section 3. 3) Section 4 describes an implemented <u>system</u> that realizes the data representation and extraction method. Section 5 describes experiments in cell manipulation using the system.

In this paper, we concentrate on the topic of behavior sampling. Some aspects of status on demand will be addressed in this paper to show the utility and applicability of behavior sampling, but the full embodiment of status on demand will be treated in a separate paper.

2. Behavior Sampling Theory

Behavior sampling entails both a data representation and a means of extracting information from the raw data to insert into the data representation format. Several concepts of behavior sampling are strongly related to the fields of hypermedia and video object representation and compression. A theoretical review of hypermedia and data models is contained in Section 2.1. The conclusion of section 2.1 discusses the extensions to hypermedia necessary to realize behavior sampling. The theoretical background of the proposed data representation is discussed in Section 2.2.

2.1. Visual Objects and Hypermedia

The primary motivation behind visually based teleoperation is the manipulation of physical objects in the slave environment through visual means. Thus a basic element would naturally be the designation of visual representations of the objects in the environment and their spatial and temporal locations in the scene. These are referred to as *visual objects*.

In addition to visual objects, there are several other channels of information in a visually based control system. The control information and relationships specified by the sensing points (see Appendix I) as well as the graphical and textual system display information must be communicated between the teleoperator and the system. In order to associate and aggregate the control information and the visual objects, it is necessary to have a data representation. If one considers the multiple associated representations of information being expressed through distinct media, "teleoperation" can be thought of as a hypermedia system. There are several useful concepts from the field of hypermedia that can be partially applied [7].

Hypermedia describes multiple media that are structured to be intra-media and inter-media navigable. Although many people are familiar with the multimedia environment provided by WWW browsers and HTML documents, those are more properly termed augmented hyper-<u>text</u> systems since it is the textual media that is intra- and inter-navigable. Although images, sounds, and videos can be included on a page and can be navigated to, elements such as sounds and video do not intra-navigate between conceptual elements in their media.

Data models for hypermedia can be categorized into three general paradigms: 1) semantic, 2) statistical, and 3) syntactical. The behavior sampled visual control data representation is based on the syntactic data model. The three paradigms will be summarized using video as an example media.

Semantic representation is the traditional human means of representing images. Visual characteristics of the image are recognized and grouped based upon abstract human-defined meanings. Typically it has been very difficult to emulate the semantic naming or cognition ability of human in machines.

The statistical viewpoint is the traditional computational processing viewpoint. Images or videos have been digitized into sequences of bit values. In statistical compression, arbitrarily grouped areas of pixels are re-assign more efficient coding symbols but the underlying structure is not exploited.

The syntactical or semiotic viewpoint exploits the underlying structure of the real world scene in the representation. Syntactic methods extract structural information without understanding the meaning or semantics of the objects since the elements can be derived through low-level vision techniques. Instead of encoding uniform blocks of images in an arbitrary manner, the spatial and temporal aspects of the scene are preserved.

Classical semiotics is based on articulatory units, termed *signs*, under the theory of "double articulation" [8]. Signs are composed of subsigns which do not have direct meaning. For example, in the text domain, words can be considered as signs composed of subsign characters. Likewise, for video image sequences which emphasize temporal differences between the images, a scene shot could be considered the sign composed of motion primitives as the subsigns.

Signs can also be combined to form *metasigns*. Thus for complex media, the various metasigns, signs and subsigns are actually a hierarchy where an element's designation as a type of sign is related to the domain and relative level of detail. For example, in video media where temporal changes in images is the dominant mode, visual objects would be considered a subsign instead of a sign. The first two columns of Table 1 show Gonzalez's proposed assignment of signs for the images and video domains [7].

Table 1. Assignment of signs for various media domains.

	DOMAIN		
	Images (Spatial)	Video (Temporal)	Control (Visual)
Meta Sign	Picture	Episode	Completed Work / assembly
Signs	Objects	Scene	individual object task
SubSigns 1	Surfaces	Shot / Global Motion	sensing pairs
SubSigns 2	Lines	Objects / Local Motion	sensing points
SubSigns 3	Pixels	Stationary Change	

Although hypermedia and object based compression do offer insight about visually based teleoperation, they are not sufficient by themselves to establish/implement extraction and accumulation of visually based control. They do not address how to: 1) extract and represent the control portion; 2) segment, isolate and extract the (manipulation) objects from the video scenes; and 3) aggregate the visual and control data.

2.2. Behavior Sampling Data Representation

Up until now, the term "media" has tended to imply the channels in human sensory communication such as aural and visual. For example, a hypermedia system for movies would encompass visual and aural information. However, behavior and control can also be communicated as well. Thus, visual control can also be semiotically described with articulation elements. For the status driven visual control domain, the last column of Table 1 introduces the sign elements for the control media proposed in this paper.

Behavior sampling entails collecting (i.e. recording) portions of both the visual domain as well as the control domain, which are closely interrelated. Thus the behavior sampling data representation must be able to encode the visual information, the control information, and the interrelationship between the control and visual information. The data representation must be structured enough to allow random or individual object access and it also should be compact enough to have feasible storage and transmission requirements.

Hypermedia representation of images and video sequences are based on the spatial or temporal relationships of the elements. These elements are typically described with nodes in a hierarchical, tree-like structure. However, some information or relationships between descendant nodes cannot be described by simple tree structures. Some lower level nodes in different spatial branches may only have direct relationships of control with each other and not other nodes in their branch. It would seem that direct cross links would be necessary to describe some relationships. Unfortunately, that leads to an inelegant solution that would be hard to manage.

One solution is to define an enhanced tree with the following three types of information in the nodes: 1) *generic*: information applicable to the node and all its descendants; 2) *intrinsic*: information applicable only to the node and not its descendants; and 3) *associative*: information that relates two or more of the node's descendants [9].

Traditional scene descriptions are able to describe the spatial temporal relationship between objects and the association of a sensing point to an object. However, a scene description alone is inappropriate to describe the overall change between the objects (as described by the control relationships between sensing point pairs). Each sensing point of the pair would be part of separate objects in the scene description and therefore would not have direct links in the description tree.

So we introduce the concept of *node control associations* (NCA) which allows explicit linkage of control information directly between arbitrary nodes through "associative information" in a higher common node. (See Figure 2) These associations can be arbitrarily added and deleted to indicate the control information changes between nodes. The most common case is the association of control information between an individual sensing point in one visual object with an individual sensing point in another visual object.

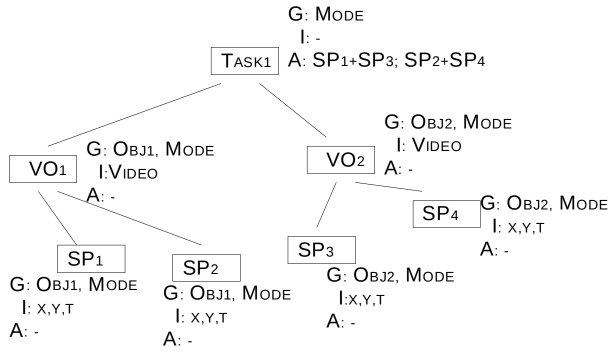


Figure 2. Visual and Control Tree. Task, Video Object (VOx), and Sensing Point (SPx) spatial information is represented in linked boxes. The adjacent Generic (G), Intrinsic (I), and Associative (A) NCA fields describe the control information.

The behavior sampling data representation includes such merits as the following: information rich record of the visual and control events; reduced storage and transmission bandwidth; multi-resolution information allows display of only the desired amount of information; multi-resolution allows priority to critical information such as sensing point locations and target object visual information and can give lower priority to background information; provides a good input for status on demand.

3. Behavior Sampling Method

3.1. Input and Output Characteristics

The input to a behavior sampling system consists of the video image of the slave environment and the time stamped control information. The control information would include such items as the location of the objects in the environment and the type of control desired. The control information is typically specified by sensing points and the desired relationship of the final state of the sensing points. All of this information will be processed and converted into the behavior sampling data representation. Further, if an operator wishes to input semantic information for a given node or link, the data representation is capable of annotating such information to the nodes and links. Semantic annotation is possible both in the initial creation phase and in post-creation phases.

The output of behavior sampling is an indexable, structured stream that contains both the visual and control information. The form of the stream is such that addressing of individual objects or information is readily obtainable without resorting to decoding all of the information in the stream or even in a large segment of the stream. The stream is suitable for storage (e.g., hard disk) or for transmission. Further, the output is parsable in such a manner that the form and style of the control performed on a given object in a past sequence is usable for control of a different object in a future situation. In other words, the behavior sampling output is suitable as the input to the status on demand functions.

3.2. Syntactical Analysis Techniques

The techniques for syntactical analysis are based upon observing the signs listed in Table 1. Syntactical analysis occurs for both the visual and control data domains. Extracted information from the two domains is complementary in that observation of a sign in one domain is often useful for structure segmentation in the other domain.

Segmentation of automated control actions is rather simplistic since the control actions occur in well-defined states and relationships. For manual operation by the operator, temporal changes are evaluated. Both the operator input events and the manipulator control system status feedback are analyzed. Events that are extracted include definition of new status points, when status points coincide, relationships between status point sets, segues, and task transitions.

The lowest level of the visual structuring begins with statistical analysis of the image sequence. Detected signs include motion of the visual objects, coincidence of visual object surface projections, global motion, and other significant changes in image statistics. Monitoring of global geometric translations and rotations allows indirect monitoring of camera work. For example, global motion can indicate a change in the operators intended work area. Likewise, analysis of localized translations and rotations indicates manipulation of the visual objects.

In visually based teleoperation, an operator is directly interested in interacting with the system in order to specify the task to be accomplished. The operator uses his knowledge to identify key points to the manipulation system though the system does not need to be aware of their underlying semantic meaning. Since the visual control itself fundamentally requires the establishment of sensing points, those also should be the basis for the visual control data representation and extraction methods.

Although the extraction of behavior from the images and status points is syntactically based, the data representation also allows semantic information to be annotated. Some semantic information can be automatically generated after syntactical analysis and using a knowledge base in manipulation situations with a priori information. Semantic information can be inferred from such sources as assumptions about the tool being used and the functions it can perform; assumptions about the objects and their roles; and operator labeling of context.

3.3. Reduction of Visual Data

During the recording phase, the preservation of all aspects of the video is typically redundant for capturing the essence of the control state changes. Although the control system needs to monitor an image tracking window surrounding each sensing point, the control system does not need to record the entire visual object. However, for later analysis by the status on demand functions, the system records the individual visual objects. Also, the recording of the visual objects is handy for human observation of the progress. Often the images of the manipulated objects themselves are sufficient for human understanding of the manipulation. The "background" both literally (in the image) and figuratively (the objects not represented by sensing points) is often unnecessary for human understanding. Thus the actual amount of imagery that is recorded and presented is often spatially and temporally sub-sampled based on its relevance to the motion behavior of the operator interacting with the manipulator. However the background information should not be completely eliminated. In some cases, humans may expect to see some type of background as there always is "background" in any natural scene, however the updating of imagery of the background can often occur only initially or rather infrequently.

3.4. Summary of Behavior Sampling

The major points of behavior sampling can be summarized as follows. It is a method for extracting image objects (fragments and full rate motion) and visual control objects (status points, etc.). Although it can be viewed as a motion picture compression technique, it is much more than a simplistic statistical technique. It can also "compress" the visual aspect of visually controlled telerobotics. When encoding control information, behavior sampling considers the inputs of the operator, changes in status points, as well as changes in the actual image. The syntactic aspect of behavior sampling itself does not directly infer the operator's task intention, however semantic information can easily be associated and annotated with the syntactic stream

4. Systems Description

4.1. System Architecture and Interface

The recording and display composer mechanisms of the first behavior sampling system are based upon the draft MPEG-4 framework [10]. MPEG-4 provides a toolbox of functions for video encoding such as specifying and encoding individual objects and specifying how the individual objects are composed to form a complete scene. MPEG-4 does not provide mechanisms for segregating or extracting objects from a video frame nor does it provide a mechanism for describing control relationships between objects.

Implementation of a behavior sampling system entailed developing two main components 1) an automated video segmentation method and 2) control information processing and storing. These are shown in the dashed boxed in Figure 3. The manipulator control section is similar in function to the SD-MHS control system. The object and scene encoding and decoding functions are part of the MPEG-4 framework.

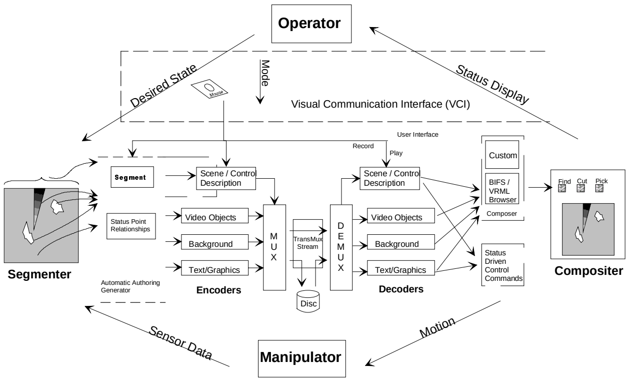


Figure 3. System Architecture

4.2. Extraction Method

Although many schemes have been introduced to automatically segment an arbitrary video frame into meaningful objects, most have encountered only limited success or required highly constrained video sequences. Further, tracking of the objects and various types of motions in the sequence is also challenging.

The behavior sampling system exploits the teleoperated control nature of the video scene to aid in the segmentation of the video image into individual control objects. When an operator initiates a teleoperation sequence, he must specify the sensing points and mode/function of the system in order to describe the final state of his manipulation desires to the system. For the recording subsystem, specification of the individual video objects is also necessary. Although edge detection and morphological techniques alone can be used to provide proposed scene segmentation based on inter pixel contrast, it is desirable to relegate as many of the proposed objects to the background object plane to dramatically reduce the number of relevant objects to be tracked and encoded. The specification of the sensing points by the operator provides an efficient and non-intrusive means of separating relevant video objects by only selecting the object edges in proximity to the sensing points.

4.3. Data Representation

Spatial-temporal hierarchical object description of the scene follows the MPEG-4 scene description mechanism: Binary Format for Scenes (BIFS)[11]. Relevant groups of pixels in the image are grouped to form the target and environment object nodes. Compound sets of object nodes form tasks. Individual objects typically will have one or more sensing points associated with it. In this aspect, the sensing points are part of the spatial-temporal aspect of the scene even though the sensing points themselves are not part of the image representation. Although MPEG-4 provides anchor points and bounding boxes to reference a given blob as a video object, those anchor points typically have no relevance to manipulation surfaces on the object. In general, more than one pixel is necessary to describe a manipulatable characteristic of a target or environment object. For example, two sensing points may be used to describe an edge for placement. Figure 4 shows portions of the textual representation of BIFS with NCA. BIFS is also encoded into a binary stream and multiplexed with the visual data stream.

```
DEF GRP Group2D {
  children [
# First Visual Object
def VO1 transform2D {
  translation 0.0 400.0    children [
      def I2 transform2D {
         translation 0.0 -0.0    children [
              image2D { url 2 } ] } ] }
              def I1002 transform2D {
                translation 20.0 38.0
                  children [
                     def SP1 SensingPoint
                          …
# Second Visual Object
def VO2 transform2D {
    …
# Update
AT 3000 {
   REPLACE V01.translation BY 100 100 }
…
# NCA
AT 4000 {
  APPEND TO VO1.children
    Content {    children [
        Generic      children[
            name{ "Object1" }
               …
```

Figure 4. Textual representation of scene description

5. Experiments in Cell Handling

5.1. System Hardware

Experiments with the behavior sampling system connected to the cell handling system (CHS) [12] were conducted. The CHS is capable of manipulating individual biological cells under optical microscopes. The scrapping tool effective width is approximately two micrometers and the translation accuracy is 0.5 micrometers.

The hardware used for the experiments includes: an Olympus BX60 microscope, 420-480 nm ultraviolet and white light sources, Sony XC-711 CCD camera, Matrox Genesis PCI image capture and processing board with TI 320C80 multi-DSP, Sigma mini-40XY pulse stepping motor stage, SMC-3(PC) pulse motor controller board, i686 MMX 266 MHz PC. (see Figure 5) The WinNT 4.0 hosted software allows mouse based two d.o.f. control of the scraper by simply clicking on the captured image of the magnified work area.

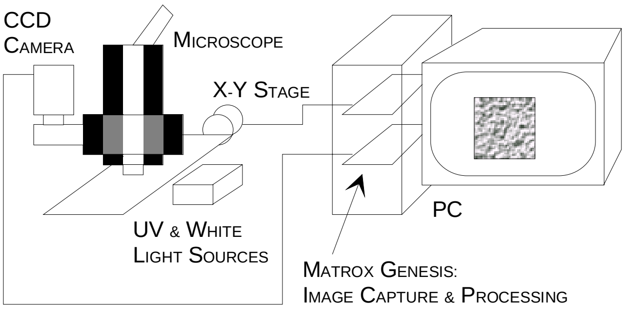


Figure 5. Block Diagram of Experiment system

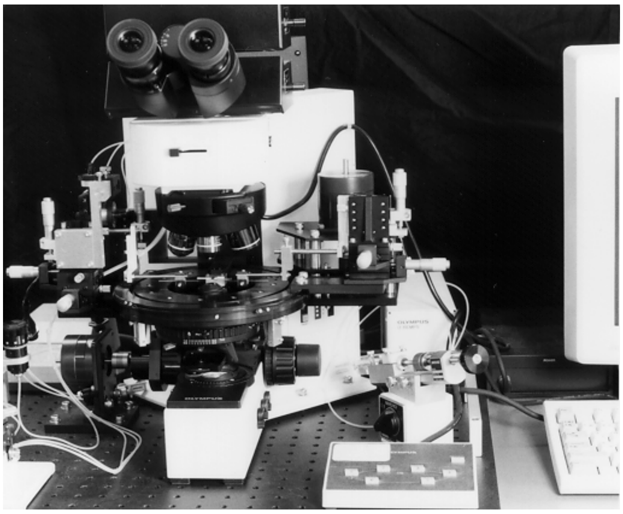
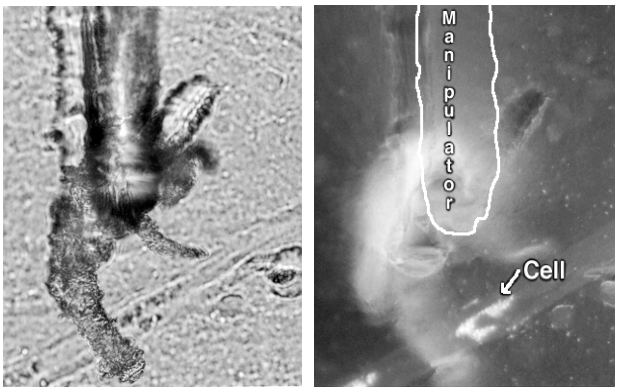


Figure 6. Microscope and Cell Handling System

5.2. Task Motivation and Description

The CHS is used in the preparation of single Mato fluorescent granular perithelial (FGP) cells for analysis. Mato FGP cells are found in the brain and studied for their relationship to aging and high-fat diets. Mato FGP cells, sometimes referred to as perivascular cells, are approximately 10 micrometers in diameter and particles inside the cell exhibit an auto-fluorescent glow in the range 520 - 570 nm (green light) when exposed to ultraviolet light. (See Figure 7(b)) This auto-fluorescent property is exploited in the image processing to help establish the approximate boundaries of the cell.



(a) (b)

Figure 7. (a) Cell illuminated with white light; (b) Cell illuminated with UV light

Individual cell analysis and manipulation is becoming increasing important for biological investigation. Heretofore methods of processing typically involved processing enmass without regard to the potentially disrupting effects of the tissue surrounding the cells. As individual cell manipulation is a developing field, there is very little human experience in such areas as how to manipulate the cell, tolerances of manipulations, appropriate tools, appropriate processes, etc. To help rapidly accumulate and exploit the new experiences and techniques currently being developed, the behavior sampling system is especially effective and being actively used. Biologists can maintain extremely accurate records of manipulation trials by visually reviewing the specific steps that a particular cell underwent. Status on demand functions can then be used to help recreate processes that were deemed effective.

One of the first examples of the Mato FGP cell processing is isolating the cell from the surrounding tissue. Although special lighting conditions combined with the auto-fluorescent property of the Mato FGP cell do provide good general cell boundary discrimination information, studies of various segmentation techniques are performed manually in order to observe the variances of the processing results. Thus behavior sampling is used to record the scraping path control information along with the visual information of the work environment.

5.3. Results

Figure 8 shows the results of an operator manually scrapping around the cell. During the cell scraping, images of the process were captured along with the corresponding control movements. The behavior sampling system combined with the cell handling system has accumulated numerous cell manipulation trials.

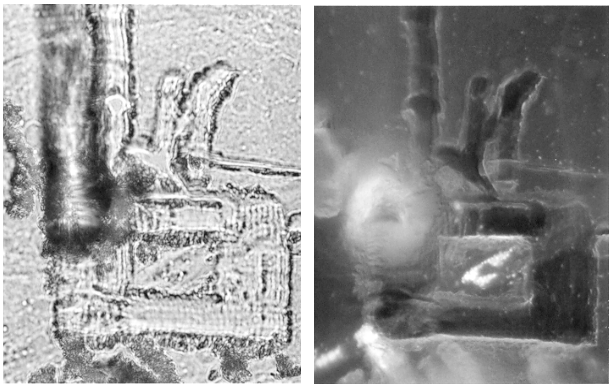


Figure 8. Results of manual scraping followed by automatic scrapping passes with increasingly wider paths.

Besides the long term analysis possible with behavior sampling, behavior sampling can be combined with status on demand to provide immediate assistance to the operator. Separating the cell from the surrounding tissue requires a band of sufficient width from the surrounding tissue. This band is wider than the scrapping area of the tool, thus several passes of the tool are required. A specialized form of the status on demand redo function has been developed to automatically scrape increasingly wider paths based upon the initial path. (See Figure 8.)

6. Conclusions

We have described a visually based teleoperation method and system which samples, structures, and stores motion control sequences and their associated imagery. This behavior sampled data can be accessed to repeat or redo a recorded sequence. Experiments using a cell handling system accumulated motion sequences of manipulating individual biological cells under optical microscopes.

Future work involves developing additional status on demand functions for general and specialized operator assistance. Finally, the status driven, behavior sampling, and status on demand components will be combined into a single system for efficient operator assistance.

Acknowledgments

We wish to thank: Takashi Miyoshi and Taichi Mita for use of the cell handling system and control software; the MPEG-4 IM1 team for their encoder and decoder software; and Kenneth Pechter and Richard Ray for insightful review of the manuscript.

Appendix I: Status Driven

Status driven is a third generation teleoperation technique where a slave manipulator is visually instructed by the master control panel. Furthering the first generation joint-angle control techniques and second generation coordinate transformation techniques, a status driven system recognizes the target status specified by the operator by extracting the task status from visual sensors.

The task environment is initially described via sensing points in the work environment and on the manipulated object. A sensing point describes a point of manipulation significance in the visual representation of the object. For example, sensing points would be used to describe the abutting surfaces in a pick and place task. The relationships between the sensing points and the change of those relationships describes the task at hand. In other words, control information is expressed in the spatial temporal relationship between corresponding sensing points associated with each object.

Operation can proceed in automatic mode where the system completely directs the slave or in shared mode where the system assists the operator by constraining the slave motion while the operator directs via a visual communication interface.

Status driven techniques are appropriate for micrometer scale part handling, the so-called “microworld assembly” where an operator’s past experience in the macro world is not applicable to the physics experienced in the microworld.

Appendix II: Status on Demand

The status on demand functionality is a visually based interface to the behavior sampled data in a status driven system. The status on demand system is able to display, through imagery, graphics, and text, major points/milestones in which (task) status has transitioned from one type of task to another. Thus, an operator is able to view the past sequence of events comprising tasks in an easy to comprehend and partition manner. The task status at each relevant point in time of the procedure is then available for reference and visual re-manipulation by the operator.

Of the several status on demand functions (e.g., redo, undo, preview, etc.) available to assist the teleoperator, only the *redo* function will be discussed here. Redo allows the operator to repeat the last style of change of state (perhaps on a different set of sensing points). This is useful with similar motions that need to be performed multiple times from (typically) different start and end points.

For example, if there is a series of objects to be manipulated in a similar way, the operator would setup the sensing points for the first object and perform one or more manipulation tasks on it. For the subsequent objects, new sensing points would be used to specify the object(s) and the Redo function would perform the same compound set of manipulations.

References

[1] S. Palm, H. Murayama, T. Mori, and T. Sato. “Visual Control through Status Driven Teleoperation”. *Advanced Robotics*, Vol. 11, No. 5, pp. 463-480, 1997.

[2] G. D. Hager, “A Modular System for Robust Positioning Using Feedback from Stereo Vision,” *IEEE Trans. On Robotics and Automation*, Vol. 13, No. 4. pp. 582-595, August 1997.

[3] N. P. Papanikolopoulos, P. K. Khosla, and T. Kanade, “Visual Tracking of a Moving Target by a Camera Mounted on a Robot: a Combination of Vision and Control,” *IEEE Trans. On Robotics and Automation*, Vol. 9, No. 1, pp. 14-35, February 1993.

[4] T. Shibata, Y. Matsumoto, and T. Kuwahara, “Hyper Scooter: a Mobile Robot Sharing Visual Information with a Human,” *Proceedings of R&A 95*, Vol. 1, pp 1074-1079, 1995.

[5] T. Sekimoto, T. Tsubouchi, S. Yuta. “A Simple Driving Device for a Vehicle - Implementation and Evaluation,” *Proceedings of IROS 97*, Vol. 1, pp. 147-154, 1997.

[6] M. Mato et al, “Involvement of Specific Macrophage-lineage Cells Surrounding Arterioles in Barrier and Scavenger Function in Brain Cortex,” *Proc. Natl. Acad. Sci. USA*, Vol. 93, pp. 3269-3274, April 1996.

[7] R. Gonzalez, “Hypermedia Data Modeling, Coding, and Semiotics,” *Proc. of the IEEE*, Vol. 85, No. 7, pp. 1111-1140, July 1997.

[8] W. Noth, "Handbook of Semiotics," Bloomington: Indiana Univ. Press, 1990, 1995.

[9] A. M. Murching et al, “Indexing Object Content Information (OCI) for MPEG-4 / MPEG-7,” ISO/IEC JTC1/SC29/WG11 M2878, Fribourg, Switzerland, October 1997.

[10] R. Koenen, “Overview of the MPEG-4 Standard,” ISO/IEC JTC1/SC29/WG11 N1730, Stockholm, July 1997.

[11] "Text for CD 14496-1 Systems," ISO/IEC JTC1/SC29/WG11 W1901, Fribourg, Switzerland, October 1997.

[12] T. Sato, T. Miyoshi, and H. Miyazaki, “Development of Cell Handling Robot System,” *Proc. Of RSJ 14th Meeting*, Vol. 3, pp. 1135-1136, November 1996. (In Japanese)