

Automatic Analysis of the Content of Cell Biological Videos and Database Organization of Their Metadata Descriptors

Andrés Rodríguez, Nicolás Guil, David M. Shotton, and Oswaldo Trelles

Abstract—We present a video content analysis and metadata organizational system for research videos arising from biological microscopy of living cells. Automated procedures are described to determine the position, size, shape and orientation of cells in each video frame. From the temporal changes in the values of these simple metadata parameters, high-level descriptors are derived that describe the semantic content of the video. This content information (*specific intrinsic metadata*) is of high information value, since it describes the behavior of cells and the timing of events within the video, including changes in environmental conditions experienced by the cells. When such metadata are properly organized in a searchable database, a content-based video query and retrieval system may be developed to locate particular objects, events or behaviors. Moreover, the availability of such semantic contents in the formal and generic format we propose will allow the application of data mining techniques and the amassing of more elaborate knowledge, e.g., species classification depending on behavior, patterns in response to environment changes, etc. The suitability and functionality of the proposed metadata model is demonstrated by the automated analysis of five different types of biological experiments, recording epithelial wound healing, bacterial multiplication, the rotations of tethered bacteria, and the swimming of motile bacteria and of human sperm.

Index Terms—Biological videos, content analysis, content recognition, semantic metadata.

I. INTRODUCTION

MOVING image data (videos, movies, animations and four-dimensional (4-D) confocal microscopy images, having the spatio-temporal dimensions of x , y (and, for 4-D confocal data, also z) and *time*, represent the most complex and demanding image type to be stored in scientific digital image databases [1]. This is both because digitized video files are typically two or three orders of magnitude larger than those for other forms of multi-dimensional image data such as three-dimensional (3-D; x , y , z) volume images [2] or multispectral satellite images [3], and also because their subsequent viewing has a time-critical component.

Manuscript received November 3, 2000; revised September 9, 2002. This work was supported in part by the European Commission under Project BIO4-CT96-0472 and under Grant 1FD97-0372 from the EU-FEDER Programme. The associate editor coordinating the review of this paper and approving it for publication was Dr. Thomas R. Gardos.

A. Rodríguez, N. Guil, and O. Trelles are with the Computer Architecture Department, University of Malaga, Campus de Teatinos, 29017 Malaga, Spain (e-mail: andresr@ac.uma.es; nico@ac.uma.es; ots@ac.uma.es).

D. M. Shotton is with the Image Bioinformatics Laboratory, Department of Zoology, University of Oxford, Oxford OX1 3PS, U.K. (e-mail: david.shotton@zoo.ox.ac.uk).

Digital Object Identifier 10.1109/TMM.2003.819581

The scientific community has only recently started to address the problem of biological video management with the development of biological image databases [4], [5]. Despite these advances, vast quantities of valuable information contained in such databases are lost due to time constraints and the lack of proper tools for automatic extraction of features and for semantic interpretation of their content (e.g., the recognition of spatio-temporal *events* involving interactions between specific content items in the video). There is thus a pressing need to develop automatic procedures for querying and recovering content-based information from this type of videos. From our perspective, it is essential to define a general formal framework to represent information about the locations of living objects within such videos, the movements of these objects, and the fundamental dependency relationships among them. Establishing a coding consensus for such biological video metadata will allow the development of new indexing, query by content and retrieval technologies.

The growth of the information culture has led to a revolution in audio-visual information systems, and it is clear that in the near future video work will be conducted entirely in the digital domain. The ability to perform efficient searches on digital moving image data requires new standards for storage, cataloguing and querying videos. MPEG-2 has been developed as a format and compression system for video broadcasting and distribution. Nowadays it is accepted as the worldwide standard for high-quality digital video compression and transmission that has been fully adopted by the video industry. In contrast, current visual information retrieval systems, based on textual information and fuzzy pattern matching, require significant improvement, as do the tools for automated feature extraction. To this end, the scientific and industrial communities are currently dedicating significant effort to the development of new standards for cataloguing and querying multimedia content [6]–[12]. Two major initiatives are involved in this effort: the Society of Motion Picture and Television Engineers (SMPTE <http://www.smpte.org>), and the Motion Picture Experts Group (MPEG; <http://www.cselt.it/mpeg>).

The MPEG-7 standard (formally known as the *Multimedia Content Description Interface*), that focuses on the description of multimedia content will provide core technology to allow the description of audiovisual data content in multimedia environments, standardizing descriptors, descriptor schemes, coding schemes, a description definition language, and system tools. *Descriptors* are representations of features that define the syntax and the semantics of each feature representation. At present, the

elements for the description of the higher conceptual aspects of video content are still under development and validation.

The SMPTE group have adopted a rather different approach, and definitive standards have yet to be published [35]. For example, the draft SMPTE Metadata Dictionary is encoded using a compact binary *key, length, value* (KLV) notation, but lacks an accompanying metadata model. Strenuous efforts are presently being made to ensure bidirectional interoperability between these two systems.

As a contribution toward the development of “high level” semantic descriptors for scientific video content, we have undertaken an interdisciplinary analysis of a specific collection of scientific video recordings, that represent a particular field of biological experimentation involving studies of the behaviors and interactions of living cells by light microscopy. Using this video collection we have identified the major types of biological video metadata, and propose methods for capturing and a model for organizing them.

The subject matter of the time-lapse and real-time video recordings chosen for this analysis include 1) the closure of in vitro wounds in epithelial cell monolayers under a variety of experimental conditions such as exposure to drugs; 2) the in vitro proliferation of *E. coli* bacterial cells; 3) the rotational movements of flagellate *Rhodobacter sphaeroides* bacteria tethered to glass coverslips using an anti-flagellin antibody, that change the direction and velocity of their flagellar rotations either spontaneously or in response to environmental stimuli; 4) the movements of free-swimming *Rhodobacter sphaeroides* bacteria in response to such flagellar rotations; and finally 5) the motility of human sperm, used to evaluate their potential ability to fertilize human eggs.

What have these experiments in common, and what factors are inherently specific for each one? How we can express the content of such videos in a compact, formal and generic manner? Is it possible to define a set of high-level *descriptors* for characters (cells), events, interactions and relationships of biological relevance?

In this work we draw answers to these challenging questions. Detailed analyses of those videos have allowed us to identify common and specific attributes for the characters (i.e., the cells) participating in the videos. The high-level descriptors we propose to describe conceptual aspects of the video content are derived from four simple character metadata: position, size, shape and orientation angle. The behaviors of the cellular characters are defined in terms of changes in these values and their evolution along the time axis.

Based on these parameters, we have defined a generic and extensible data model that provides a coherent description of the multimedia content of such videos. These metadata are then properly organized in a searchable database that supports subsequent queries to locate particular characters, events or behaviors, that may be correlated with changes of environmental conditions occurring during the recordings.

The suitability and functionality of the proposed metadata model is demonstrated by means of the automatic analysis of the different experiments. This analysis has involved 1) the identification of the discrete objects in each image sequence, and the tracking of the movement of these objects in space and time,

using image and video processing techniques, 2) the application of new *image understanding and event analysis procedures* that permit the automatic detection of specific behaviors and interactions, 3) the development of an automatic event annotation procedure used to populate a suitably organized video metadata database, and finally 4) the accessing and querying of resulting video metadata database. Preliminary reports of this work have been presented at recent international conferences [13]–[15].

II. SYSTEM AND METHODS

The proposed strategy for video analysis and content-based querying can be formulated in the following steps (see Fig. 1):

Image processing and object detection: Initially, image processing is required to identify the discrete objects (cells, “characters”) in each image sequence, and to track the movement of these objects along the space/time axis. To this end, various visual feature extraction algorithms have been combined and improved, in the light of detailed biological knowledge of the systems.

Image understanding and metadata annotation: Using the data from the trajectories of the individual objects, we address a particular issue of image understanding, namely the automatic detection of events. After event detection, an automatic event annotation procedure is used to populate the video metadata database in accordance with the organizational structure of the entity-relationship model described as follows.

Query formulation and information retrieval: The spatio-temporal attributes of the objects and events detected or defined in the previous steps are subsequently used for the video query and retrieval system. The result of a successful query is the identification of a particular video sequence or a set of video clips from among those stored in the database.

A. Metadata Definition

We have defined and distinguished three classes of metadata relating to images and video recordings [16]:

Ancillary metadata are simple text-based details about an image or video, concerning, for example, the filename, title, format, subject and author, and technical details of the image or video preparation, but which do *not* relate specifically and directly to the visual content of the image or video frames. In the MPEG-7 terminology, these elements, listed in [10], describe the content from the *Creation & Production, Media, and Usage* viewpoints. In most cases, these ancillary metadata are *generated* by the author of the image or video, and cannot be extracted from the image content.

Generic intrinsic metadata relate to attributes of images and video frames such as color, size, shape, texture and (in the case of video) motion. These are reducible to numerical image primitives than, once extracted, may be employed to describe the content from the MPEG-7 *Structural Aspects* viewpoint [10] or to locate images using the simple query by example paradigm “*Find me any image or video frame that looks like this one.*” Such metadata may also be used to produce a storyboard for a video (transitions, scene segmentation, etc.), or to create the foundation for a frame-accurate index that provides nonlinear

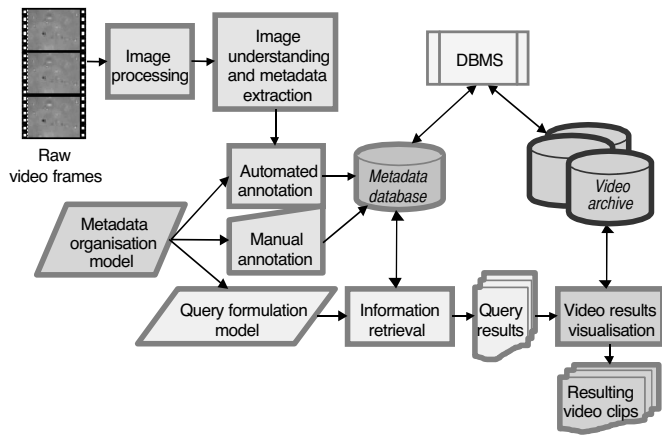


Fig. 1. *System Overview*. Initial image processing and image understanding procedures supported by biological knowledge allows the automated identification of objects and events (changes in the behavioral states of individuals and/or groups; interactions between content items), which may be supplemented by manual annotation, generating specific intrinsic metadata that are organized using a flexible and extensible data model and stored in a relational database. Subsequent queries result in the retrieval of video clips matching the search criteria.

access to the video [17]. The use of such generic intrinsic metadata for searching image and video libraries has in recent years become the subject of intense research, with a large and expanding bibliography (see, for example, [32], [33]), and proprietary software for this purpose is now available from a number of companies (e.g., [34]).

Specific intrinsic metadata, resulting from intelligent manual or automated analysis of the images or video frames, describe the spatial positions of specific objects within images, and the spatio-temporal locations of objects and events within videos. These metadata correspond with the MPEG-7 *Conceptual Aspects* viewpoint of the audiovisual content.

Of the three metadata types, we have focused our work on this latter one, which is the most rewarding in terms of information content, since it relates directly to the spatio-temporal features within a video that are of most immediate importance to our human understanding of video content, namely “*Where, when and why is what happening to whom?*” Subsequent query by content on this kind of metadata extends the query domain from the conventional one of textual keyword or image matching techniques to include direct interrogation of the spatio-temporal attributes of the objects of real interest within the video, and of their associated event information.

B. Metadata Model

The core of the proposed metadata organization is the definition of three main types of entities: *Media Entities*, *Content Items* and *Events* [16]. The specific intrinsic metadata describing the video content according to this schema are stored using the data model presented in Fig. 2. It is worth while to highlight that for interoperability, the proposed metadata model could be easily ported to SMPTE or MPEG-7 formats by using the appropriate SMPTE metadata dictionary definitions, encoded using the key-length-value (KLV) code [18], and, when available, the W3C XML schema used to encode MPEG-7 descriptors.

1) *Media Entities*: Cell biological videos may be primary recordings of experimental observations, consisting simply of a single shot recorded as a continuous video sequence by a single camera with a fixed field of view, or they may be more complex, consisting of different shots and scenes edited together for a particular educational or research purpose. Definition of the media entities *video*, *scene*, *shot*, and *frame* is essential for the subsequent definition of the content items and events contained within them.

As in conventional video analysis, we divide videos into scenes, each of which relates to a different aspect of the total video, and subdivide scenes into shots, each of which is a single contiguous series of video frames derived from one camera take. While some of their parameters, for example the time-lapse ratio (i.e., the ratio between frame recording rate and frame playback rate), are strictly *ancillary* metadata that cannot be derived from analysis of the video itself, they are included here for convenience. Other important ancillary metadata items relating to the media entities (e.g., video format, Digital Object Identifier [19], [20]; www.doi.org, and INDECS rights metadata [21], [22]; www.indecs.org are not relevant to our present analyses, and are intentionally omitted from the following description.

2) *Content Items*: Content items include animate *characters* (e.g., cells) and inanimate *objects* (e.g., micropipettes) appearing in the video frames that share many properties in common. They are recorded in separate content item tables for convenience, and may each be organized into groups and classes. For example, individual bacterial cells may be divided into two groups, “tethered” and “free,” both members of the class “*Rhodobacter sphaeroides* bacteria.” For simplicity in the subsequent text of this paper, only characters are discussed, but the metadata organization for objects is identical.

3) *Events*: An *event* is an instantaneous or temporally-extended action or “happening” that may involve a single character or object (e.g., a cell contracting), an interaction between two or more characters (e.g., a cytotoxic cell killing a target cell), or all the characters (e.g., perfusion of the observation chamber by a drug). It is helpful to group events into *event categories* to aid subsequent searching.

C. Metadata Generation

1) *Measurement of Primary Spatio-Temporal Parameters*: The primary *specific intrinsic metadata* relating to individual content items are, in our analyses, generated using traditional, enhanced or, in several cases, specific image processing techniques used to segment individual frames, from which the sequential spatial positions of the moving content items may be determined from frame to frame. For each content item that is correctly identified, the following primary spatial parameters are recorded for each video frame: position, size, shape and orientation. These values may then be used to describe the complex behavior of the content items, as described at the end of this section. Because of the huge amount of content items (bacteria) in videos, the validation of the item identification has been performed automatically, based on the matching between the observed item features

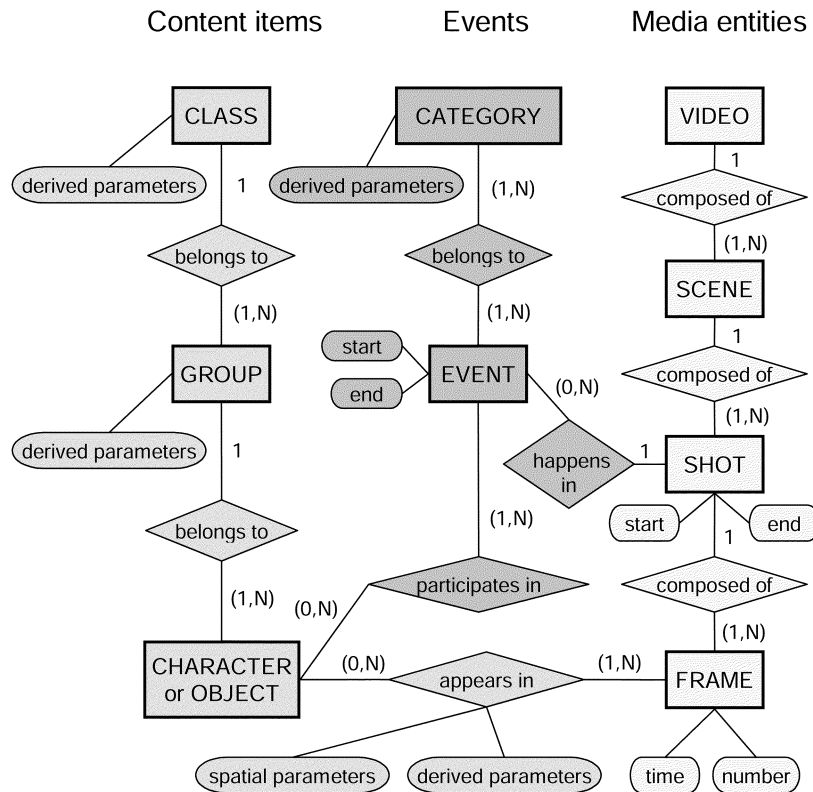


Fig. 2. The *metadata entity-relationship model*. The model shows the relationships between media entities, content items and events [16]. To preserve simplicity and clarity within this two-dimensional diagram, certain relationships (e.g., between groups and frames) have not been represented. In addition, the starts and ends of videos and scenes have not been shown explicitly, and the derived parameters of groups, classes and event categories are shown associated with each of these items, rather than with their relationships to media entities, as should strictly be the case. The notations on the lines joining the items show their entity relationships (e.g., (1)–(1,N) means “one to one or one to many”).

and the statistically expected ones (appropriate size and shape, continuity in space position, etc.).

2) *Calculation of Derived Parameters*: Given the positions, sizes, shapes and orientations of all the content items in every frame (the primary spatio-temporal specific intrinsic metadata), derived specific intrinsic metadata may be determined, for example, the rate of growth of a particular cell, or the handedness of its rotation.

For motile cells, tracking procedures are applied to the primary spatio-temporal metadata in order to obtain the trajectory characteristics (motion history) of each content item, which includes the start and end frames of a particular track, and the instantaneous velocity from frame to frame.

In addition to these instantaneous parameters, one can also compute average parameters over longer time periods, for example the average translational or rotational velocity, direction of movement or growth rate of a single cell, or the proportion of time spent in a particular state (e. g. stationary or swimming), either for the entire duration of the recording of a single identified character, or over a defined period (e.g., the last 50 frames). From these data one can also determine the variances of these parameters.

Furthermore, from the parameters relating to individual characters or objects, derived specific intrinsic metadata describing the *population behavior* of an entire group or class of content items may also be obtained. Here it is necessary to define what we understand to be the appropriate metadata in the context of a

population. In most cases, they will be the *average* values from all the component members of a group or class (mean size, average speed, etc.), together with the variance of the parameter and a record of the number of individuals comprising the group. However, in others cases alternative descriptors are appropriate. For example, in studying bacterial growth, the group size (i.e., the number of cells forming the entire colony) is the metric of primary importance (see Section III-B). Similarly, we may wish to record the group position, defined as the centre of gravity of the area occupied by the population, and the degree of dispersion or clustering within the population. We may also wish to compute some probabilistic information, such as the probability of occurrence of a given event within the population.

These derived parameters, examples of which are given in Fig. 3, may be predefined and recorded in the metadata database, or may be computed on the fly from the primary metadata as and when they are required. We have adopted the strategy of storing the four primary *spatio-temporal parameters* (low-level descriptors) described above, for each object in every frame together with their corresponding derived parameters. In this way we expect to obtain all the advantages of maintaining low-level information (new derived parameters can be computed when necessary), and fast access and processing based on high-level information.

It is worth emphasising that while the primary spatial parameters are generic for all type of videos, and closely following the proposed MPEG-7 and SMPTE descriptors, descriptors

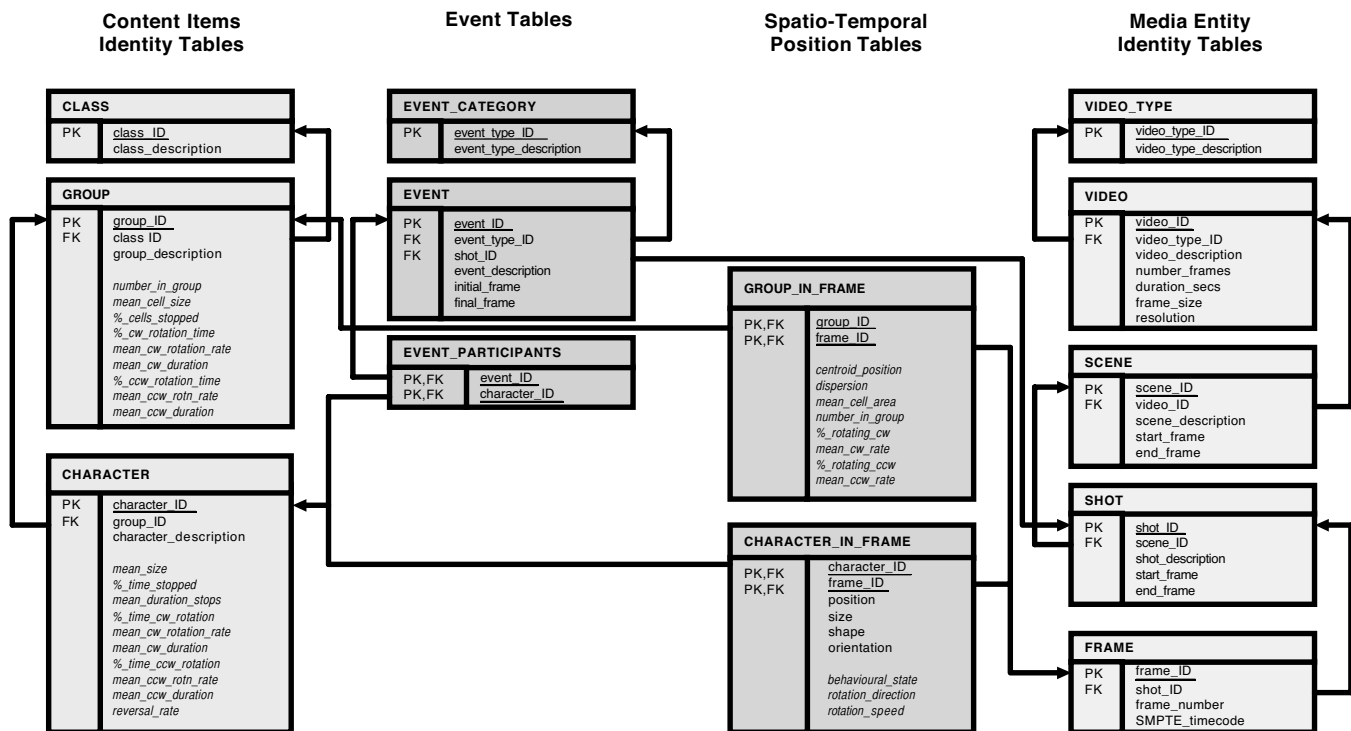


Fig. 3. Relational tables for storage of the specific intrinsic metadata (rotating bacteria case). Content item identity tables, events tables, spatio-temporal position tables and media entity identity tables are shown from left to right, using the same general layout and color scheme as in the metadata entity-relationship model (Fig. 2). Metadata parameters common to all videos are shown in normal font. For illustrative purposes, derived metadata parameters relevant to the analysis of videos of rotating bacteria (detailed in Section III-D) are shown within the tables in italics. For simplicity, the content item identity table for objects (as opposed to characters), and the spatio-temporal position table for *Class_in_Frame*, are not shown. These can be inserted as and when required, and follow exactly the hierarchical format of the tables shown. Most ancillary metadata (e.g., author, video format, color depth/dynamic range, and time lapse ratio) are also not shown.

within these standards are still not available for many of these higher-level derived behavioral parameters based upon them (derived specific intrinsic metadata).

3) *Identification of Activity States and Events:* Next, we address the most *application dependent* activity of the analysis, namely that concerned with identifying the *activity states* of each character, and the *events* in which it is involved. For this, specific biological knowledge is required to define the activity states and events of biological relevance in each particular experiment. In some cases, it will be important to record the movements of the object, for example changes in velocity in response to a change in an environmental variable such as temperature, turns in otherwise linear trajectories, or transitions between swimming and stationary, or between clockwise and counter-clockwise rotation. In other analyses, it will be important to determine the timing of significant life events such as cell division or cell death.

The derived specific intrinsic metadata required for different types of video analyses will vary. However, the metadata model permits specific intrinsic metadata relating to novel video items to be added easily, so that the same database system can be used for the analysis of a wide range of different types of videos [13]–[15]. The spatio-temporal parameters of the characters, objects and events thus determined, properly organized in a searchable database (see Fig. 3), allow subsequent queries to locate particular characters or events.

At present, situations that involve more than one cell type, such as the infection of a host cell by a parasite, or the induction

of apoptotic cell death in a target cell by a cytotoxic effector cell [23], are too complex for fully automated video analysis. For such situations, we have developed a prototype alternative system named *VANQUIS* (video analysis and query interface system) which permits content analysis to be conducted in a semi-automated manner, using the superior pattern recognition skills of the human eye-brain system to permit the user to make the decisions, while employing the computer to speed the tasks of metadata annotation, database entry and subsequent query by content [24].

III. ANALYSES AND RESULTS

The application domain chosen for our demonstration of video analysis for subsequent query by content, using the specific intrinsic metadata described in the previous section, involves the analysis of scientific videos of cell biological specimens. Five types of biological videos have been used as the model subjects for this initial work, all derived from recordings of living cells observed using the light microscope.

A. Wound Healing Videos

The *wound healing videos* analyzed were made using time-lapse video microscopy with phase contrast optics to record the closure of in vitro wounds made in freshly confluent monolayers of cultured epithelial cells under a variety of experimental conditions, for example the transient perfusion exposure of the cells with drugs that effect cytoskeletal polymerization [25]–[27]. For

this type of video, questions of interest relate to the rates of wound healing, as measured by reduction of the cell free area, as healing proceeds under different drug regimes, in contrast, for example, to the rate of growth of the same type of cells at the free margins of unwounded colonies.

The automatic feature extraction procedure for this type of videos is designed to measure the rate of wound healing by the progressive loss of open wound area (see Fig. 4). To this end, we have used computer image processing techniques to distinguish the uniform character of the cell-free open wound in each frame from those regions of cellular monolayer on either side, where the image texture is quite different, being characterized by high frequency modulations of the image intensity.

Segmentation of the uniform wound region from the rest of the image is undertaken by using a differential operator that calculates the changes in the first spatial derivative of the image [28]. Further histogram equalization and image thresholding allows us to exclude high contrast zones. The resultant threshold closely follows the faint margins of the ruffling wound edge cells.

Once the image processing has been completed, an internal bounding box is used to compute, for each video frame, the wound area per unit length of wound, and the edge lengths. This internal bounding box, which is used as a safeguard to exclude end effects, is defined as a rectangle oriented parallel with the axis of the linear wound (computed by principal components), which lies fully within the video frame, and which is wider than the maximum wound width. The size and position of this bounding box (see Fig. 4) is kept constant for all the frames of the video that are to be analyzed.

The wound area within the bounding box is then treated as a single object of interest in the identity table, and no metadata are recorded concerning the positions and shapes of the individual cells. The change of this wound area per unit length of wound per minute permits the rate of wound closure, and hence the effectiveness of the healing process, to be calculated [14], [15]. Significant changes in the healing rate may be related to changes in the culture environment of the cells.

B. Bacteria Proliferation Videos

The second analysis is closely similar to the first one, in that both measure the loss of cell-free area in successive video frames. Here, the *in vitro* proliferation of *E. coli* bacterial cells is recorded by time-lapse video microscopy using Nomarski optics (Fig. 5). Given favorable culture conditions, bacteria grow rapidly and undergo mitotic cell division approximately every half hour, each division resulting in the generation of two daughter cells from the original cell. When analysing this type of video, the objective is to determine the rates of cell proliferation at different stages in the culture.

The automatic feature extraction procedures used in this case are very similar to those used to measure wound healing in the first example, clearly demonstrating the reusability of image processing software designed to this end. Visual inspection of the frames to be processed suggested the use of a texture based algorithm to solve the problem of image segmentation, since

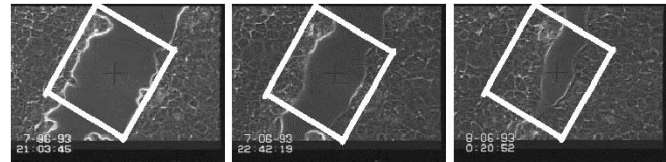


Fig. 4. Stages during *in vitro* epithelial cell wound healing. A gallery of video frames recorded at three points during one of the wound healing videos. Times are given as hh:mm:ss in the lower left corner of each frame, below the original date of the recording. The width of each video frame is approximately 300 μm . A clear and distinguishable difference in texture can be observed between the uniform character of the cell-free open wound and the regions of cellular monolayer on either side.

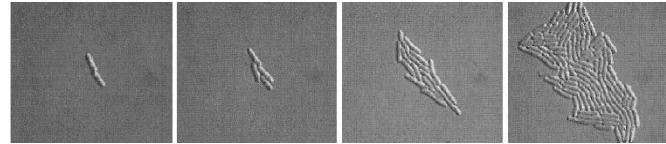


Fig. 5. Stages in the early growth of an *E. coli* culture. An image sequence showing successive time-points during bacterial proliferation. The width of each video frame is approximately 20 μm , and the time interval between images is approximately 30 min.

the background lacks the high spatial frequency information present in the region occupied by the growing cells. Use of first derivative information and a region growing procedure used to close small gaps in the cell region, allows the identification and grouping of nonuniform regions into a single area occupied by the cell colony. The outer perimeter of the colony defines an area that is then used as a measure that approximates the number of cells in the colony at each moment. Occasionally, uniform cell-free regions are enclosed within the colony perimeter, leading to abrupt and artificial increases in the estimated cell number. In these cases, a more precise method based on edge detection of the individual cells is applied. This method both identifies the edges of the population and eliminates any uniform internal regions.

Metadata from analyses of the wound healing and bacterial proliferation videos: Using the specific intrinsic metadata for the cell-free areas derived from frame by frame analyses (Figs. 4 and 5), the instantaneous rates of wound healing and of bacterial culture growth can easily be computed (Fig. 6), and from these the mean rates of wound healing or colony growth can be determined over any desired time period, or in response to drug treatment.

In the wound healing study, for example, this is achieved by looking at the appropriate register in the spatio-temporal position table, where the open wound area is recorded for each frame in the video, with its associated derived metadata, e.g., the instantaneous wound healing rate. These parameters are dynamic features of the content character *wound*. So they must be stored in our metadata database in the spatio-temporal table *CHARACTER_IN_FRAME* represented in Fig. 3, that records the dynamic semantic state of every character throughout the video. Changes in this rate in response to drug administration may be determined by searching the database for differences in the healing rate before and after the event of drug administration, the timing of which is recorded in the events table.

C. Bacterial Motility Videos

This set of analyses are of videos recording in real time (a) the movements of free-swimming *Rhodobacter sphaeroides* bacteria as their flagellar rotations change velocity or direction, and (b) the rotations of another population of the same species of bacteria that have been tethered to the glass coverslip of the microscope observation chamber using an anti-flagellin antibody, both made in the laboratory of Professor Judy Armitage of the University of Oxford [29]. For these types of video, questions of interest relate to the determination of individual and population statistics concerning run lengths and velocities, and the frequencies, durations and patterns of tumbles of free-swimming bacteria, and concerning the rotational directions and speeds of tethered bacteria, and the frequencies and patterns of their stops and reversals, correlated with changes of environmental conditions.

The bacterial motility videos contain large numbers of “characters” (the bacteria) whose movements are independent from one another, presenting a high level of complexity for the analysis and metadata extraction. The commercial system presently in use by the creators of the videos for the analysis of bacterial motility [30] has severe limitations in the number of bacteria that can be simultaneously tracked, and in the extent of the data that can be analyzed and stored, both problems related to the fact that it is designed to work with limited hardware resources in real time direct from an analogue video camera or a videotape.

In a first stage of our own offline analyses of these videos [13], [15], an initial segmentation of the frame images is undertaken with due regard for the variations in background illumination between frames, using a dynamic thresholding procedure [31]. Subsequently, the sizes and shapes of the individual bacteria are determined using a growing region algorithm, where bacterial “objects” are built from an initial seed point inside each bacterium. The orientation of the main axes of the elongated cells is determined by principal component analysis. For each cell, we can thus record its initial position, size, shape and orientation in space [Fig. 7(a)].

For the videos of the free-swimming bacteria, the next step is to track the movements of the cells [Fig. 7(b)]. The tracking problem can be defined as one of recognising the same object in consecutive frames of the video. The initial algorithm used to solve this problem is simple, and relies on the fact that any bacterium is likely to show a similar size, shape, and orientation in adjacent frames of the video, and that its position in any frame is likely to be close to that in the preceding frame. Application of this algorithm results in the determination of bacterial trajectories from which parameters such as speed, direction and curvature may be calculated.

However, since the individual bacteria are swimming unrestricted in three dimensions in the space between the microscope slide and the overlying coverslip, they may stray from the narrow focal plane of the microscope objective lens and become temporarily lost from view. This causes them to be missed by the initial segmentation and cell recognition algorithms, causing fragmentation of their trajectories. Since for the scientific analysis of bacterial movement it is important to have

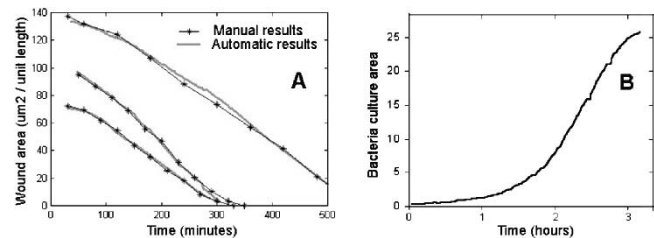


Fig. 6. Graphs showing rates of wound healing and bacterial growth. (a) Plots of the data obtained by fully automated analysis of changes in wound area with time for three separate wound healing experiments in which drug treatment was not applied. The data is shown after filtering to remove noise caused by periodic dark frames in the original video (see [14] for explanation), superimposed on the results obtained in 1994 by manual analysis of the original analogue video recordings [25]. (b) The growth curve of a culture of *E. coli* cells obtained by fully automated analysis, showing an initial lag phase while the colony proliferates slowly from one to 500 cells, followed by a period of exponential growth from 500 cells to 17000 cells (the so-called log phase, which gives a linear plot when the logarithm of the cell number is plotted against time), and ending in a gradual reduction in growth rate during the colony’s multiplication from 17000 to 25000 cells, as more and more bacteria compete for dwindling nutrient resources.

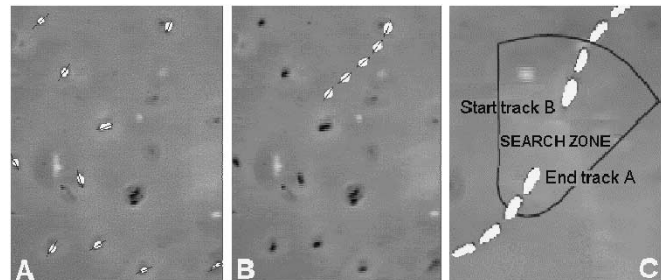


Fig. 7. Automated identification and tracking of swimming bacteria. In a first stage of analysis, segmentation of the images is performed to detect individual bacteria in each frame. (a) All the single bacteria in the focal plane have been successfully detected and are shown in white, outlined, and with a principle axis vector drawn. Those bacteria that are clustered in small groups, where recognition of individual cells is difficult, and those bacteria above or below the focal plane, which appear as indistinct light or dark blobs, have intentionally been excluded from the selection. (b) The trajectories of the bacteria are determined. The result of this procedure applied to a single bacterium is shown in the central image, where the bacterium’s positions in five consecutive frames are shown superimposed upon an image of the initial frame. (c) A schematic representation of the strategy used to solve the incomplete trajectory problem. The two different tracks detected, which show bacterial positions at single frame intervals, are actually sections of the same bacterial trajectory, broken when it swam out of the plane of focus. For truncated tracks such as these, the algorithm tries to find corresponding truncated tracks and link them.

trajectories that are as long as possible, there is a need to link partial or broken trajectories into longer continuous ones. This is achieved by a postprocessing algorithm that checks, for every partial trajectory that ends, whether there is another partial trajectory which is spatially adjacent, which starts within an appropriate time interval after the ending of the first trajectory (a few frames later), and which matches the first one in features such as speed and direction, and the shape and size of the bacterium. If these conditions are fulfilled, the “two” bacteria originally identified are recognized to be the same cell, and the two trajectories are linked to form a longer one [see Fig. 7(c)]

For the free swimming bacteria, the important high-level events to detect are the transitions between their three principle behavioral states [Fig. 8(a)] namely the *forward swimming*

state when all the bacterial flagella are rotating counter-clockwise which is characterized by a smooth continuous curvilinear movement, the *tumbling* state, in which the clockwise rotation of the flagella cause the bacterium to gyrate randomly; and the *stationary* state. This last state is the easiest to define, and is used to categorise all bacteria with translational velocities less than a very low threshold value. To distinguish the forward swimming state from the tumbling state, we use a sliding time window of w frames length (typically 8–10 frames), over which we compute the ratio between the total distance travelled (i.e., the sum of all the interframe distances) and the overall displacement (i.e., the linear distance between the initial and final positions). Values near to one characterize forward swimming, while higher values represent the tumbling state.

For each identified bacterium, the system determines and stores the specific intrinsic metadata relating to the cell's spatio-temporal trajectory. From these primary metadata one can calculate derived parameters such as the instantaneous velocity, the duration, initial direction and curvature of the individual forward swimming trajectories, and the frequency, duration and patterns of tumbles and stops. Fig. 8(b) shows a typical example of bacterial tracking where five tumbles have been detected, interrupting forward swimming and causing random changes in direction. The motility parameters can be correlated with details about the environmental conditions pertaining at the time. In addition, summary statistical metadata may be produced describing the average motility of the whole bacterial population in the video.

For the rotating tethered bacteria, the task of identifying the same cell in successive video frames is obviously more straightforward, and the salient features to record from such videos are the instantaneous speed, handedness and duration of each rotation, accelerations and decelerations, the frequency of reversals, and the duration of stops. These metadata are then stored in the appropriate identity, spatio-temporal position and events tables, as detailed in Fig. 3. Results from such analyses are not shown here, but are presented in [13, Fig. 5].

Examples of queries that can be made using the specific intrinsic metadata derived from videos of the free swimming bacteria are: "Identify all the video clips containing any bacteria swimming at mean individual velocities of at least x μm per second, and "Find me the video sequences in which, after the administration of drug A, the average tumble frequency decreases by more than 30%." For the first query, a simple inspection of the appropriate spatio-temporal position table permits identification of the video frames containing bacteria with individual velocities, averaged over the preceding 25 frames (1 s), above x μm per second. The second question requires a calculation of the average tumble frequency of the population of bacteria before and after the event of administration of drug A, determined from the times of the starts of all bacterial tumbles and of drug administration recorded in the events table.

D. Human Sperm Motility Videos

Real time recordings of human sperm motility are widely used to assess the potential ability of human semen to fertilize a human egg, there being a good correlation between sperm movement characteristics and fertilization success. Several

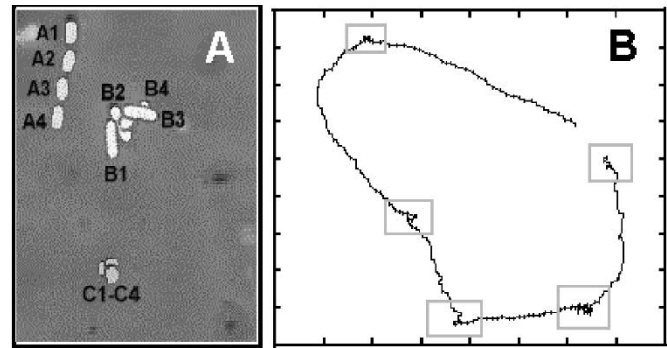


Fig. 8. Behavior states and event detection in bacterial swimming. (a) The positions of three different bacteria are shown in four successive frames. By analysing the lineal distance covered, the directional change in trajectory, and the change in cellular orientation during a brief interval, it is possible to determine the current behavioral state of each bacterium. Bacterium A is swimming, bacterium B is tumbling, and bacterium C is stationary. A degree of flexibility (percentage hysteresis) is applied to the thresholds that separate these states, in order to avoid 'noisy' behaviors of cells close to a threshold. (b) The trajectory of a single bacterium plotting over ten seconds, in which the system has automatically detected five tumbles (marked with boxes).

characteristics are involved in studies of this type, including the sperm "count" (i.e., the total cell density, typically required to be greater than 20 million per ml for fertility) and the "motile density" (i.e., the density of motile sperm, typically required to be greater than 8 million motile cells per ml). The motile density is perhaps the most important parameter, as it reflects the number of sperm capable of progressing through the female reproductive tract from the site of deposition near the cervix to the site of fertilization in the Fallopian tube.

For this type of video recording, questions of interest relate to changes in mean swimming velocity of the sperm, and the proportion of motile sperm, determined at hourly intervals over a period of up to 24 hours from receipt of the specimen. It is well known that sperm motility is temperature dependent, so the handling and processing of specimens is critical, requiring that specimens be evaluated only in tightly controlled laboratory conditions.

The automatic feature extraction procedures used in this case are very similar to those used to measure bacterial motility (Section III-C). Both types of video contain large number of characters that shown a variety of behaviors. For these videos, questions of interest relate to the determination of individual and population statistics concerning run lengths and velocities, and the frequencies, durations and patterns of circular and linear movements measured by the rotational directions and linear speeds of spermatozooids.

Here we present (due to length constraints) the analysis of a circular trajectory movement (see Fig. 9). After an initial segmentation of the frame images that allows the identification of individual spermatozooids (position, size, shape and orientation of the main axes), a tracking procedure is applied to determine the trajectories, as is explained in Section III-C.

IV. DISCUSSION

In this work, we have devised a general framework for the analysis and storage of metadata describing the content of cell biological research videos, and have employed a variety of

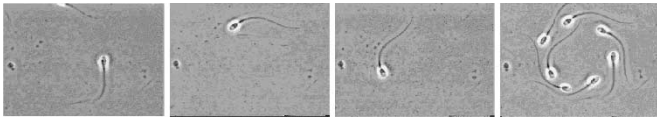


Fig. 9. Phase contrast images of human sperm. Successive video frames are shown on the left (sampled at ten frames), illustrating the motility of a single human sperm cell in vitro. The cell is approximately $70\ \mu\text{m}$ long. The right image, which is a superposition of seven frames, shows the unusual circular trajectory of this particular cell, which is rotating once every 1.62 s.

algorithms for feature extraction from the digitalized videos. Based on the proposed metadata model, we have developed an automatic analysis and annotation system for this type of scientific video recording.

We have named this prototype analytical system *VAMPORT* (Video Automated Metadata Production by Object Recognition and Tracking). It permits the generation of information of high intrinsic value by intelligent analysis of the visual content of moving image data. In all cases, the results we have obtained have been validated by alternative conventional methods of analysis, and we have observed significant increases in both accuracy and efficiency when undertaking such analyses automatically rather than by hand.

For the studies of epithelial wound healing and bacterial growth, useful quantitative data have been obtained simply by determining the change in the cell-free area with time. For our analysis of the motility of individual cells, we have determined primary specific intrinsic metadata relating to four related spatial parameters: the position, size, shape and orientation of each identified cell in each video frame. These metadata descriptors closely follow the approved recommendations of the MPEG-7 standard for representing information about video content.

From these primary spatial metadata, and their variations along the time axis of the video, we have computed a variety of high-level features related to its semantic content described by a set of derived parameters. Furthermore, from detailed knowledge of their biology, we have been able to define the potential behavioral states of particular cells, permitting the automated identification of transition events between these states. All of these primary and derived specific intrinsic metadata parameters relating to individual cells can be combined or averaged to determine group characteristics.

The specific intrinsic metadata extracted in this way are stored and organized in a relational database, over which a content-based query and retrieval prototype has been built. Subsequent queries on these metadata may be used to obtain important factual and analytical information, and to retrieve selected video sequences matching the query criteria. As result of a successful query, the system returns to the user a list of video files, together with details of the relevant frame numbers, allowing video clips matching the query to be retrieved.

As an extension of this study, the specific intrinsic metadata stored in the database may be subjected to further analysis and mining algorithms to produce more elaborate knowledge, for example the discovery of common patterns in the motility behavior of two species of bacteria, or the classification of a single bacterial population into two classes (e.g., normal and

nontumbling mutants). In the future, such data mining will be increasingly important in many areas of biomedical and pharmaceutical research, but can only be truly successful and widespread if the initial analyses and metadata encodings are all made to conform to recognized international standards, such as MPEG-7 and the SMPTE metadata dictionary, to ensure true interoperability between separate data sets and databases. Since, in the framework of MPEG-7 and SMPTE developments, high-level specific intrinsic metadata descriptors such as we have described are still under definition and validation, the data model and the descriptors that we propose here may be of value in refining the emerging definition of these important standards.

ACKNOWLEDGMENT

The authors are most grateful to Prof. J. Armitage of the Bacteriology Unit, Department of Biochemistry, University of Oxford, and to Dr. J. Sullivan of Cell Alive (<http://www.cell-salive.com>), for permitting us to use as test data for our analyses their video recordings of bacterial motility, and of sperm motility and bacterial proliferation, respectively, and to Dr. Philippe Salembier of the Image Processing Group, Universidad Politecnica de Catalunya for valuable comments on this manuscript.

REFERENCES

- [1] T. Boudier and D. M. Shotton, "Video on the internet: an introduction to the digital encoding, compression and transmission of moving image data," *J. Struct. Biol.*, vol. 125, pp. 133–155, 1999.
- [2] J. L. Sussman, D. Lin, J. Jiang, N. O. Manning, J. Prilusky, O. Ritter, and E. E. Abola, "Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules," *Acta Crystall. D.*, vol. 54, pp. 1078–1084, 1998.
- [3] National Centre for Geographic Information, Ministerio de Fomento (2000). [Online]. Available: <http://www.cnig.ign.es>
- [4] J. M. Carazo and E. H. K. Stelzer, "The BioImage Database Project: organizing multi-dimensional biological images in an object-relational database," *J. Struct. Biol.*, vol. 155, pp. 97–102, 1999.
- [5] The Global Image Database, GlaxoWellcome Experimental Research Group. [Online]. Available: <http://www.gwer.ch/qv/gid/gid.htm>
- [6] F. Nack and A. T. Lindsay, "Everything you wanted to know about MPEG-7: part I," *IEEE Multimedia*, vol. 6, pp. 65–77, 1999.
- [7] J. Hunter, DDL Working Draft 1.0, 1999.
- [8] A. Lindsay, MPEG-7 Applicat. Doc. v.9, 1999.
- [9] Overview of the MPEG-7 Standard, J. M. Martinez, Ed., 2000.
- [10] P. van Beek, A. B. Benitez, J. Heur, J. Martinez, P. Salembier, J. Smith, and T. Walker, MPEG-7 Multimedia Description Schemes XM (version 2.0), 2000.
- [11] P. Salembier, "Visual segmented tree creation for MPEG-7 description schemes," in *IEEE Int. Conf. on Multimedia and Expo ICME*, July 18, 2000.
- [12] —, "Audiovisual content description and retrieval: tools and MPEG-7 standardization activities," in *Tutorial: IEEE Int. Conf. on Image Processing, ICIP-2000*, Vancouver, BC, Canada, Sept. 10, 2000.
- [13] A. Rodriguez, D. M. Shotton, N. Guil, and O. Trelles, "Automatic tracking of moving bacterial cells in scientific videos," in *Proc. RIAO 2000, 6th Conference on Content-Based Multimedia Information Access*. Paris: Coll. France, Apr. 2000.
- [14] A. Rodriguez, D. M. Shotton, O. Trelles, and N. Guil, "Automatic feature extraction in wound healing videos," in *Proc. RIAO 2000, 6th Conference on Content-Based Multimedia Information Access*. Paris: Coll. France, Apr. 2000.
- [15] D. M. Shotton, A. Rodriguez, N. Guil, and O. Trelles, "Analysis and content-based querying of biological microscopy videos," in *Proc. ICPR2000, 15th Int. Conf. on Pattern Recognition*, Barcelona, Spain, Sept. 4–7, 2000.

- [16] —, "A metadata classification schema for semantic content analysis of videos," *J. Microscopy*, pt. 1, vol. 205, pp. 33–42, Jan. 2002.
- [17] A. Gupta, *Visual Information Retrieval Technology—A Virage Perspective*: Virage, Inc. White Paper, 1997.
- [18] Proposed SMPTE Standard SMPTE 336M. Data Encoding Protocol Using Key-Length-Value.
- [19] N. Paskin, "Toward unique identifiers," *Proc. IEEE*, vol. 87, pp. 1208–1227, 1999.
- [20] —, (2000) The DOI Handbook. [Online] Available http://www.doi.org/handbook_2000/index.html.
- [21] G. Rust. (1998, July) Metadata: The Right Approach. D-Lib Magazine. [Online]. Available: <http://www.dlib.org/dlib/july98/rust/07rust.html>
- [22] G. Rust and M. Bide. (1999) Introduction to the INDECS Metadata Schema. [Online]. Available: <http://www.indecs.org>
- [23] D. M. Shotton and A. Attaran, "Variant antigenic peptide promotes cytotoxic T lymphocyte adhesion to target cells without cytotoxicity," *Proc. Nat. Acad. Sci.*, vol. 95, pp. 15 571–15 576, 1998.
- [24] J. Machtynger and D. M. Shotton, "VANQUIS, a system for interactive semantic content analysis and subsequent query by content of videos," *J. Microscopy*, vol. 205, pp. 43–52, 2002.
- [25] J. W. Lewis, "The Effects of Colchicine and Brefeldin A on the Rate of Experimental In Vitro Epithelial Wound Healing," Undergraduate Research Project Dissertation, Univ. Oxford, U.K., 1994.
- [26] J. W. Lewis and D. M. Shotton, "Effects of colchicine and brefeldin A on the rate of experimental in vitro epithelial wound healing," in *Proc. 4th Annu. Meeting Eur. Tissue Repair Soc.*, Oxford, U.K., 1994, p. 230.
- [27] —, "Time-lapse video microscopy of wound healing in epithelial cell monolayers: effects of drugs that induce microtubule depolymerization and Golgi disruption," *Proc. Roy. Microscopy Soc.*, vol. 30, pp. 1–135, 1995.
- [28] N. Guil, J. M. Gonzalez, and E. L. Zapata, "Bidimensional shape detection using an invariant approach," *J. Pattern Recogniti.*, vol. 32, pp. 1025–1038, 1999.
- [29] M. Manson, J. Armitage, J. Hoch, and R. Macnab, "Bacterial locomotion and signal transduction," *J. Bacteriology*, vol. 180, pp. 1009–1022, 1998.
- [30] G. Hobson, *Hobson Tracker User Manual*. Sheffield S11 9ND, U.K.: Hobson Tracking Systems Ltd., 1996.
- [31] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982, pp. 143–146.
- [32] Indexing, storage, browsing, and retrieval of images and video, in Special Issue of *J. Vis. Commun. Image Represent.*, S. Panchanathan and B. Liu, Eds., vol. 7, pp. 305–423, 1996, doi:10.1006/jvci.1996.0026.
- [33] Content based access of images and video libraries, in Special Issue of *J. Vis. Commun. Image Represent.*, A. Del Bimbo, V. Castelli, S.-F. Chang, and C.-S. Li, Eds., vol. 75, pp. 1–213, 1999, doi:10.1006/cviu.1999.0775.
- [34] A. Hampapur, A. Gupta, B. Horowitz, S. Chiao Fe, C. Fuller, J. Bach, M. Gorkani, and R. Jain, "The Virage video engine," *Proc. SPIE*, vol. 3022, pp. 188–198, 1997.
- [35] J. Hunter and L. Armstrong, "A comparison of schemas for video metadata representation," *Comput. Networks*, vol. 31, pp. 1431–1451, 1999.



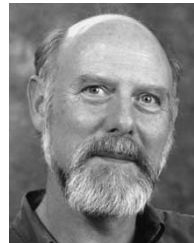
Andrés Rodríguez received the M.Sc. and Ph.D. degrees in computer science from the University of Málaga, Spain, in 1995 and 2000, respectively.

Since 1996, he is an Assistant Professor with the Department of Computer Architecture, University of Málaga. His research interests include video processing and video content description, data mining and query by content in multimedia databases.



Nicolás Guil received the B.S. degree in physics from the University of Sevilla, Spain, in 1986 and the Ph.D. degree in computer science from the University of Málaga, Spain, in 1995.

From 1990 to 1997, he was an Assistant Professor at the University of Málaga. Currently, he is an Associate Professor with the Department of Computer Architecture, University of Málaga. His research interests are in the area of video and image processing: tools for video indexing, automatic and semi-automatic video segmentation, image retrieval, and active contours.



David M. Shotton received the B.S., M.S., and Ph.D. degrees from the University of Cambridge during the late 1960s.

He subsequently worked at Bristol University, U.K., the University of California, Berkeley, Harvard University, Cambridge, MA, the Weizmann Institute of Science, Israel, and Imperial College London, U.K., before joining the Biological Sciences Faculty of the University of Oxford, U.K., in 1981, where he is now a University Lecturer in cell biology within the Department of Zoology. His cell biological research primarily involves the use of time-lapse video microscopy and freeze-fracture electron microscopy to investigate cellular function. He has published extensively on light and electron microscopy techniques, and has taught frequently on international microscopy courses. He is a founding partner of the original BioImage Database consortium, and recently established the Image Bioinformatics Laboratory within the Department of Zoology to handle and analyze digital images and videos of biological specimens. He is presently director of the BioImage Database development (www.bioimage.org) within the EC 5th Framework ORIEL Project (www.oriel.org), and also director of VideoWorks for the Grid, a new U.K. e-Science Testbed Project (www.video-works.ac.uk) within which VIDOS (www.vidos.ac.uk), a Web-based video editing and customization service, is being developed. His current research interests include semantic content description and query-by-content of scientific videos.



Oswaldo Trelles received the B.S. degree in industrial engineering from the Universidad Católica de Perú, the M.S. degree in computer sciences from the Polytechnical University of Madrid, Spain, and the Ph.D. degree, working on parallelization of biological sequences analysis algorithms on supercomputers, from the University of Málaga, Spain.

Currently, he is an Associate Professor, Department of Computer Architecture, University of Málaga, Spain, where he is teaching fundamental and design of operating systems with the Computer Science faculty. His main interest research areas include parallel paradigms for distributed and shared memory parallel computers and graphical support for exploratory analysis. Data mining and automatic knowledge discovering in biological data are also of great interest in his research.