Annotation of Cultural Heritage Documents Based on XML Dictionaries and Data Clustering

Zenonas Theodosiou[1], Olga Georgiou[1], Nicolas Tsapatsoulis[1], Anastasis Kounoudes[2], and Marios Milis[2]

[1] Dept. of Communication and Internet studies, Cyprus University of Technology, 31 Arch.Kyprianos, P.O. Box 50329, 3603, Limassol, Cyprus
{zenonas.theodosiou,olga.georgiou, nicolas.tsapatsoulis}@cut.ac.cy
[2] SignalGeneriX Ltd, Arch.Leontiou A' Maximos Court B', 3rd floor, P.O. Box 51341, 3504, Limassol, Cyprus
{tasos,marios.milis}@signalgenerix.com

Abstract. Cultural heritage forms the local and national identities. It shapes relationships between neighbors and other communities around the world. The sweet wine named “Commandaria” is part of Cypriot heritage and currently holds a protected destination of origin within European Union, USA and Canada. In the framework of the Commandaria project we managed to gather an enormous amount of data, related to Commandaria wine, corresponding to photographs, scanned documents and videos. The need of a method for efficient retrieval of these data based on their actual content was mandatory. The data were appropriately indexed through a multilevel labeling scheme allowing access from various modalities and for a variety of applications. Despite the huge efforts for automatic characterization and classification human intervention is the only way for reliable multimedia data annotation. Manual data annotation is an extremely laborious process and efficient tools developed for this purpose can make, in many cases, the true difference. In this paper we present the CulHIAT, a cultural heritage item annotation tool, which uses structured knowledge, in the form of XML dictionaries, combined with a hierarchical classification scheme, to attach semantic labels to image and video segments at various levels of granularity. Finally, XML dictionary creation and editing tools are available during annotation allowing the user to always use the semantic label she/he wishes instead of the automatically created ones.

Keywords: Cultural Heritage Documents, Video Annotation, Hierarchical Classification, XML Dictionaries, Data Clustering.

1 Introduction

Cultural heritage is the legacy of physical artifacts and intangible attributes of a group or society that are inherited from past generations, maintained in the present and bestowed for the benefit of future generations. The heritage that survives from the past is often unique and irreplaceable. Commandaria wine, is for many years one of most important Cypriot cultural heritage items. In the context of the Commandaria project

M. Ioannides (Ed.): EuroMed 2010, LNCS 6436, pp. 306–317, 2010.

[1] huge amounts of literature work have been scanned, true stories, relics, tools and related items were recorder or animated, and agriculture photographs were captured to provide better access to researchers in the area of Humanities, to primary and secondary school teachers, to university students and the great public in general.

No more than twenty years ago we faced the problem of information unavailability or inaccessibility; today we face issues related to data overflow and information filtering, which make users' investigation for relevant information cumbersome [2]. The growing amount of digital video and images are driving the need for more effective methods for indexing, searching, and retrieving of video and images based on their content. While recent advances in content analysis, feature extraction, and classification are improving capabilities for effectively searching and filtering digital video content, the process to reliably and efficiently index multimedia data is still a challenging issue. Besides, in order to learn audio-visual concept models, supervised learning machines also require ground truth labels being associated with training videos [3].

In consistency with the broader digital multimedia domain, manual annotation is the predominant way of attaching description and indexing information to digitized cultural heritage items [4]. Manual annotation, however, is a laborious, time consuming and expensive task. Furthermore, the terminology and description of cultural heritage are often too technical and difficult for nonprofessional users of the domain [5]. Thus, annotation based on a dictionary according to the different knowledge levels of users is a critical aspect of annotation enhancement. The use of an integrated system which can provide the dictionary creation and data annotation can greatly simplify the process.

In this paper we present the CulHIAT cultural heritage item annotation tool which is based on MPEG-7 standard [6], [7], [8]. It supports image, video and digital data annotation though a highly interactive panel. Video annotation at lower levels of granularity is supported through automatic video segmentation into shots while shots are combined into scenes through clustering with the aid of a Genetic Algorithm scheme. A new XML dictionary based on the Commandaria taxonomy was created to attach semantic labels to an enormous amount of images, videos and other digitized or digital documents related to the Commandaria wine with the aid of this tool. Although many annotation tools are publicly available, few of them can be used to annotate data in multiple levels of granularity as required, for example, for audiovisual documents.

Annotation can be performed in several manners, ranging from completely manual to tool-assisted to fully automatic. *VideoAnnEx* MPEG-7 annotation tool is implemented by IBM for collaborative multimedia annotation task in distributed environment [3]. *MovieTool* is developed by Ricoh for creating video content descriptions conforming to MPEG-7 syntax interactively [9]. *Multimodal Annotation Tool* is a release of IBM's Multimedia Mining Project, which is derived from an earlier version of VideoAnnEx with special features with audio signal graphs and manual audio segmentation functions [10]. *Microsoft Research Annotation System (MRAS),* which is a web-based system for annotating multimedia web content [11]. Synchronous *Multimedia and Annotation Tool (SMAT)* is used to annotate images and there is no granularity for video annotations nor controlled-term labels [12]. *A4SM (Authoring System for Syntactic, Semantic and Semiotic Modelling)* developed a semi-automated annotation tool for audio-visual media in news [13]. *The European Cultural heritage*

Online (ECHO) is developing a multimedia annotation tools which allows people to work collaboratively on a resource and to add comments to it [14].

While the majority of the above tools are very well-developed and efficient they emphasize on a particular medium, usually video. Furthermore, the annotation lexicons that are used to attach labels to the data are predefined and non-editable. Modifications to these lexicons or creation of new ones cannot be done in "context" (during the annotation process) and require the use of external tools. This is clearly a limitation since the annotator rarely stops the annotation process to modify the lexicons (even if we assume that he/she is allowed to do so). Lexicon enrichment, on the other hand, is an important facility and must be preserved because it leads to more effective data indexing and data retrieval.

The remaining of this paper is organized as follows: Section 2 gives a brief description of the Annotation Tool, while the dictionary creation and annotation procedure are described in Section 3. In Section 4 we present the algorithm that is used for shot clustering. Finally, conclusions are drawn and further work hints are given in Section 5.

2 CuLHIAT Annotation Tool

The CulHIAT annotation tool was developed based on MPEG-7 standard using MATLAB. The tool boasts a user-friendly Graphical User Interface allowing the management of multimedia content (images, videos or any other digital or digitized document), video segmentation into shots and scenes, video and image annotation, image watermarking and creation of XML dictionaries. Any PC or workstation can be used as far as the hardware is concerned, although as expected, the more computing power the better performance level it will reached. In particular, a large amount of RAM memory will help to improve the performance of video segmentation and classification.

The GUI consists of three major panels. The *Video Panel* provides the video or animation segmentation and annotation while the *Image Panel* provides the image (or scanned documents) annotation and watermarking. The dictionaries used for annotation are created via the third panel named *Dictionary Panel*. A brief description of each panel follows.

2.1 Video Panel

An example screen of *Video Panel* is shown in Fig. 1. The Panel consists of three tabs. The first and second ones provide the manual and automated segmentation respectively, while the third accords the video annotation.

Video Segmentation: Video segmentation is required to split a video sequence into smaller video units. As shown in Fig. 1, the user can select one of two available modes for video segmentation: manual and automated. The video sequence displayed in the window on the upper left-hand corner during manual segmentation. The user can explore it and set the shots boundaries by specifying the first and last frame of each shot. Then she/he might choose and set key-frames. Key-frame are representative images of the video shot, and offer an instantaneous recap of the whole video shot. The shot frame boundaries and key-frames of each shot along with any other annotation information are saved in an XML-file related with the particular video sequence.

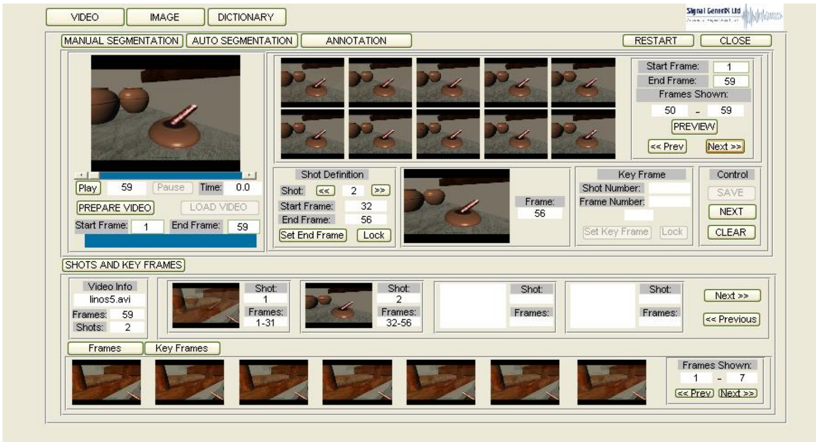

Fig. 1. Video Segmentation panel. The frames shown present a segment from an animated video clip created to present a production phase of Commandaria. It represents real Commandaria production procedure at LINOS, a renovated wine press located at Lania village in Cyprus.

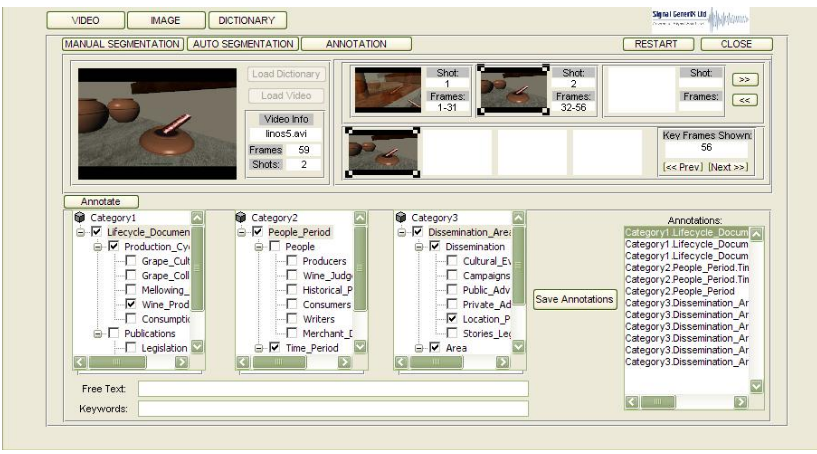

Fig. 2. Video Annotation panel. This figure presents the annotation of a frame shot, taken from Commandaria production phase, video. The dictionary panel was defined based on Commandaria taxonomy structure.

Automated shot detection is also supported with the aid of the Color Histogram Differences algorithm [15]. The algorithm is one of the most trustworthy variants of histogram-based detection algorithms and is based on the idea that the color content rapidly changes across shots. So, hard cuts and other short-lasting transitions can be detected as single peaks in the time series of the differences between color histograms of contiguous frames or of frames a certain distance *k* apart.

Let $p_i(r, g, b)$ be the number of pixels of color (r, g, b) in frame I_i of N pixels. Each color component is discretized to 2^B different values, resulting in $r, g, b \in [0, 2^B-1]$.

Usually B is set to 2 or 3 in order to reduce sensitivity to noise and slight light object as well as view changes. Then the color histogram difference CHD_i between two color frames I_{i-1} *and* I_i *is* given by:

$$CHD_i = \frac{1}{N} \cdot \sum_{r=0}^{2^B-1} \sum_{g=0}^{2^B-1} \sum_{b=0}^{2^B-1} \left| p_i(r,g,b) - p_{i-1}(r,g,b) \right| \tag{1}$$

If within a local environment of radius I_c of frame I_i only CHD_i exceeds a certain threshold, then a hard cut is detected. As presented in [15], for particular type of hard cut which consists of one transitional frame, in a pre-processing stage double peaks (i.e. groups of $S_c=2$) contiguous CHD_i exceeding threshold were modified into single peaks at the higher CHD_i.

A video clip can simply annotated by describing its content in its entirety. However, when the video sequence is large, indexing particular video segments is invaluable. Smaller video units might correspond to video shots or a collection of similar video shots usually referred to as scene. Shot detection has already been described earlier while scene construction is performed using shot clustering as described in Section 4.

Given the shot boundaries, the annotations are assigned for each video shot by using the *Video Annotation* Tab. The tool uses a specific type of dictionaries based on MPEG-7 descriptions made via the *Dictionary Panel*. To be more precise the video annotation is performed through the following three steps.

Video Annotation: First, the annotation dictionary (XML) that will be used to annotate the key-frames of each shot is loaded. Annotation dictionaries can be created using the *Dictionary Panel* as will be explained in Section 2.3. Each dictionary can be divided up to three main categories, as illustrated in Fig. 2, usually corresponding to different annotation contexts or annotation perspectives. This choice was made for space fitting purposes since every main category can be further divided to any number of subcategories and any subcategory to any number of nodes and so on, providing a hierarchical annotation tree.

Second, the segmented video resulted from the video segmentation procedure is loaded and its shots are shown in shot axes at the upper right-hand corner. After choosing a shot, its key-frames are shown in the four axes below the shot axes. The user can choose a key-frame in order to annotate it. The chosen key-frame can be seen in the axes at the left-hand corner.

Third, the key-frame annotation can be implemented using the dictionary categories presented in list-boxes on the bottom left corner of the video annotation panel. The user ticks the boxes of the most representative labels and adds if needed free text and key-words using the corresponding edit boxes. Annotations are shown in the list-box on the bottom right corner and are saved into an XML file in the video directory.

2.2 Image Panel

CulHIAT provides the capability of annotating and embedding information into an image, scanned document or digital item via the *Image Panel*. A screen shot of the *Image Panel* is shown in Fig. 3. The input image and its features are presented on the top left corner of the panel. The user can choose between *Image Annotation* and *Watermarking*.

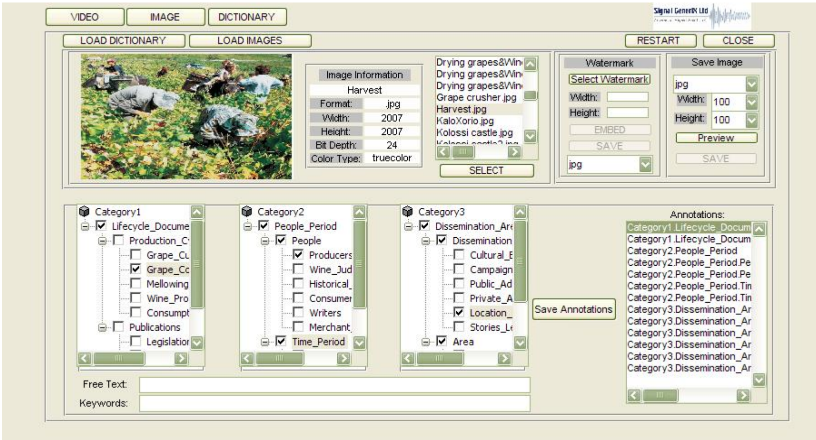

Fig. 3. Image Annotation. The image shown presents the grape collection procedure during grape collection season, which is usually starts early September and ends late October.

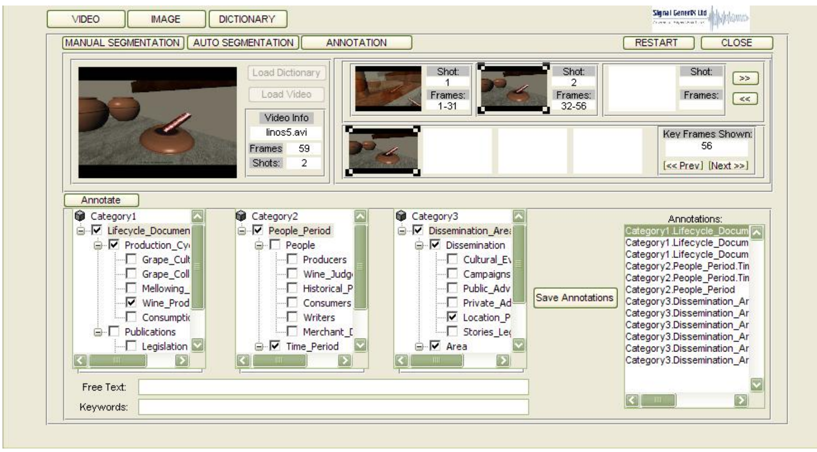

Fig. 4. Dictionary Panel. The shown dictionary was created based on Commandaria taxonomy and involves six main sections.

Image Annotation: An image can be simply annotated following the three steps described in the previous section. The three categories of the annotation dictionary are presented in the list-boxes (under left-hand) and the chosen image is shown at the upper left-hand axes. The user can annotate the image by ticking the most appropriate annotation boxes of the three lists. Free text and keywords edit boxes can be used for a more detailed annotation. The saved annotations are shown in the list-box on the right-hand corner and are saved into an XML file.

Image Watermarking: The selected image can be watermarked and saved in any image format and in any scaling using the *Watermark* tab. After the watermark selection, the user defines the wanted width and height. The watermarked image can be seen in the left-corner axes and can be saved in any image format. The chosen image also can be rescaled and saved in any format, without being watermarked using the *Save Image* tab.

2.3 Dictionary Panel

Dictionaries used for annotation via the CulHIAT are created using *Dictionary Panel*. As mentioned earlier each dictionary consists of three different categories as presented in Fig. 4. Categories can be divided in subcategories, subcategories into nodes, nodes into subnodes, etc, providing an hierarchical structure. The user can create a new dictionary from scratch or add/delete a node at any level in the hierarchy. During the annotation process any number of node-labels at any level can be used to characterize the data. However the selection of a node-label implies also the selection of its ancestor node-labels. Dictionaries are saved in an XML file.

3 Cultural Heritage Data Annotation

An enormous amount of data was collected in the framework of Commandaria project. Some of these data date back Homer's era indicating the historical value of Commandaria wine. Proper annotation of collected data items was an absolute requirement in order to allow efficient information retrieval related to Commandaria for various categories of users including students, tourists, scientists, wine producers, etc.

The Commandaria data collection consists approximately of 7500 files. The 3500 files are digitized manuscripts and scanned papers from books, journals and official legislating documents, while the remaining 4000 files are images and videos. Most of the images are taken using a digital camera, from visits towards the Cyprus agriculture and well known wine villages. Video files correspond to interviews with experienced wine producers as well as recordings of Commandaria production cycle steps. Additionally, project archives consist of short clips presenting agriculture tools and places, used for wine production. Some of the Commandaria production cycles were modeled using 2D and 3D animations.

Currently the size of data collection is around 35 GB but it continuously grows. The value of the collected information is priceless for Cyprus heritage, therefore the collection, proper preservation and easy access to this information is a task of tremendous importance [1]. Efficient indexing and retrieval of this information requires accurate and rich annotation of data; this annotation was done with the aid of CulHIAT. For example, an important evidence discovered during the data collection process, is the interview with a resident of the Cypriot village Kornos (Fig. 5(a)) who found at his grandfather's house (Fig. 5(b)) a substance in almost solid form which is probably Commandaria in concentrated form (Fig. 5(c)-5(i)). The condensed substance, which, as reported by the resident, dates more than one hundred years, was generated from a natural evaporation of alcohol and water. The Commandaria existed for many years in wooden barrels until 1976, when the majority was sold and a small part was transferred to glass bottles, preserved until today.

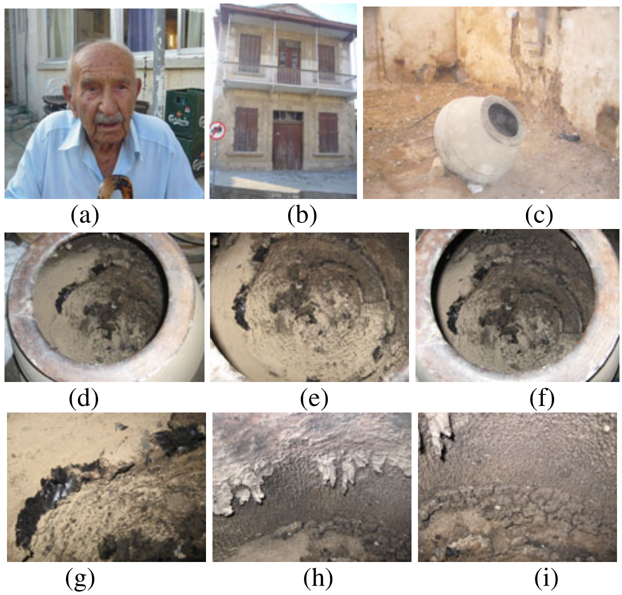

Fig. 5. Photographs taken from the Cypriot village Kornos during the interview with the resident found the 100 years old wine. (a) Mr. Sergides is the person that owns the old wine, (b) family house where the solid wine found. (c) The traditional wine making tool, so called "Pithari", which stored the wine for so many years. (d), (e), (f) present the internal content, taken from different angles. (g), (h), (i) show the solid situation of the wine, after so many years kept in a big jar.

It is clear that the observing the images shown in Fig. 5, the overall story and the important of this evidence cannot be understood. Proper annotation of these images allows not only the classification and indexing according to the Commandaria devised taxonomy [1] but also the attachment of free text description which enables and deeper appreciation of the discovered item.

As mentioned earlier, the CulHIAT annotation tool was used to characterize the Commandaria documents. For this purpose a new XML dictionary based on the Commandaria taxonomy was created using the tool, as described in section 2.3. An example of this dictionary is shown in Fig. 6.

The dictionary includes all the possible sources of the collected data, the users or producers of these data, the related historical period, production cycle or any other related information. In the production cycle we annotate images related to the production of Commandaria, starting from the grapes cultivation wine consumption. A large number of images were related to wine producers' interviews, places of Commandaria region and people working during the Commandaria production. All the collected data fall into a specific time period, therefore a relevant category was created to help us separate the corresponding information based on the historical period they belong to.

```
- <root>
  - <Lifecycle_Documentation>
    - <Production_Cycle>
        <Grape_Cultivation />
        <Grape_Collection />
        <Mellowing_Draining />
        <Wine_Production />
        <Consumption />
      </Production_Cycle>
    - <Publications>
        <Legislation />
        <Books />
        <Research />
        <Wine_Review_Results />
      </Publications>
    </Lifecycle_Documentation>
  - <People_Period>
    - <People>
        <Producers />
        <Wine_Judges />
        <Historical_People />
        <Consumers />
        <Writers />
        <Merchant_Dealer_Trader />
      </People>
    - <Time_Period>
        <Ancient_Times />
        <Middle_Times />
        <Modern_Times />
      </Time_Period>
    </People_Period>
```

Fig. 6. An XML dictionary example

4 Shot Clustering Using Genetic Algorithm

In this section we present a genetic algorithm which is used for shot clustering in order to create scenes. Scenes are composed of conceptually similar shots. Shots belonging to the same scene must be abutted each other. The algorithm assumes that the input data is divided into a desired number N_c of clusters. We assume also that the input data is already annotated with aid of method presented in the previous section.

We define the set **K** = $\{K_1, K_2,..., K_{N_k}\}$ of data points with a data point K_i corresponding to cluster C_i. Let us define a vector of integer values in increasing order:

$$Idx^{(i)} = \left\{Idx_1^{(i)}, Idx_2^{(i)},..., Idx_{N_{c-1}}^{(i)}\right\} \tag{2}$$

with $1 < Idx_j^{(i)} < N_k$, $j = 1,..., N_{c-1}$. Each vector $Idx^{(i)}$ defines a partition $P^{(i)} = \left\{P_1^{(i)}, P_2^{(i)},...P_{N_c}^{(i)}\right\}$, of set **K** with $P_j^{(i)}$ corresponding to a set of data points $\{K_{Idx_{j-1}^{(i)}},...K_{Idx_j^{(i)}-1}\}$, while $P_1^{(i)} = \left\{K_1^{(i)},..., K_{Idx_1^{(i)}-1}\right\}$ and $P_{N_c}^{(i)} = \left\{K^{(i)}_{Idx^{(i)}_{N_c-1}},..., K_{N_K}\right\}$. Given that each data point corresponds to a cluster, the partition $P^{(i)}$ defines a possible division of input data into N_c clusters. The task of the genetic algorithm described next is to find a partition $P^{(\xi)}$ which creates the optimum division of input data into N_c clusters given a properly defined metric.

4.1 A Genetic Algorithm for Shot Clustering

Genetic Algorithms are adaptive optimization methods that resemble the evolution mechanisms of biological species [16]. Feature selection is one of the areas that GAs present excellent performance. The main advantages of GAs are:

- They do not require the continuity of parameter space and,
- They are able to efficiently search over a wide range of parameters /parameter sets.

In a GA, the search begins from a population of P_N possible solutions (in our case strings corresponding to integer vectors $Idx^{(i)}$, $i = 1,..., P_N$ of length $N_c - 1$, with integer values limited to the interval [1 N_k], and not just one possible solution. Solution refers to a partition $P^{(i)}$ as explained in the previous section. A population of solutions guarantees that search will not be trapped in a local optimum, especially if significant diversity exists among the various solutions. The population of solutions tends to evolve toward increasingly better regions of the search space through the use of certain randomized processes, called genetic operators. Typical genetic operators are the *selection*, *mutation* and *recombination*. The *selection* process chooses strings with better objective function value and reproduces them more often than their counterparts with worse objective function value. Thus, a new population is formed consisting of the strings that perform better in their environment. The *recombination* (crossover) operator allows for the mixing of parental information, which is then passed to their descendants. The initial population is randomly acquired; this means that the first and major degree of diversity is introduced in this stage of the GA. The second and lesser degree of diversity is introduced when *mutation* operator acts upon each string of the population. The whole evolution process stops after a predefined maximum number of iterations (generations) is reached or the variation among population of solutions is too small.

Once the initial population has been created the process of creating new generations starts and consists, typically, of three stages:

- A fitness value (measure of optimality) of each string in the random population is calculated.
- Genetic operators, corresponding to mathematical models of simple laws of nature, like reproduction, crossover and mutation are applied to the population and result in the creation of a new population.
- The new population replaces the old one.

In our case the fitness function F is a metric of similarity between data points corresponding to the same cluster divided by the similarity of data points corresponding to different clusters. Equation 3 gives the mathematical notation of the fitness function F_i corresponding to the string $Idx^{(i)}$ ($\|\ .\ \|$ refers to the second norm of a multidimensional matrix):

$$F_i = \sum_{j=1}^{N_s} \left(\frac{\sum_{K_l, K_m \in P_j^{(i)}, l \neq m} \|K_l - K_m\|}{\sum_{K_l \in P_k^i, k \neq j} \|K_l - K_p\|} \right) \tag{3}$$

The objective is to find the string that maximizes the fitness function F. The realization of the genetic operators (reproduction, mutation and crossover) is as follows:

Reproduction. The fitness function F is used in the classical “roulette” wheel reproduction operator that gives higher probability of reproduction to the strings with better fitness according to the following procedure:

1. An order number, say q, is assigned to the population strings. That is q ranges from 1 to P_N, where P_N is the size of population.
2. The sum of fitness values (F_{sum}) of all strings in the population is calculated.
3. The interval [0 F_{sum}] is divided into P_N sub-intervals each of one being [SF_{q-1} SF_q] where

$$SF_{q-1} = \sum_{i=1}^{q-1} F_i, \quad q > 1 \tag{4}$$

$$(SF_{q-1} = 0 \text{ for } q = 0 \text{ and } q = 1)$$

$$SF_q = \sum_{i=1}^{q} F_i, \quad \forall q \tag{5}$$

F_i is the value of fitness function for the $i-th$ string (see equation 3).

4. A random real number R_0 lying in the interval [0 F_{sum}] is selected.
5. The string having the same order number as the subinterval of R_0 is selected.
6. Steps (4) and (5) are repeated P_N times in order to produce the intermediate population to which the other genetic operators will be applied.

Crossover. Given two strings (parents) of length N_{c-1} an integer number $1 < r < N_{c-1}$ is randomly selected. The two strings retain their gene values up to gene r and interchange the values of the remaining genes creating two new strings (offspring). Obviously the integer numbers in offspring must be reordered so as to correspond to vectors of integer values in increasing order.

Mutation. This operator is applied to each gene of a string and it alters its content, with a small probability. The mutation operator is actually a random number that is selected and depending on whether it exceeds a predefined limit it changes the value of a gene. If gene r is to be mutated, the allowable values $Idx_r^{(i)}$ for it are those in the interval ($Idx_{r-1}^{(i)}$ $Idx_{r+1}^{(i)}$).

5 Conclusions and Further Work

In this paper we presented an annotation tool for Cultural Heritage documents named CulHIAT. CulHIAT can be used for annotation of digital or digitized data, including video and animations, at various levels of granularity. XML dictionaries are used as predefined lexicons while the tool provides an easy to use interface to modify these or create new dictionaries. An enormous amount of data related to the sweet Cypriot wine Commandaria was annotated using this tool with the aid of newly created XML dictionary based on the Commandaria taxonomy.

CulHIAT allows for semantic labeling using structured knowledge in the form of XML dictionaries and provides a powerful algorithm for shot detection algorithm which minimizes the human intervention for video segmentation at lowest level. In addition, with the aid of an intelligent clustering algorithm shots are grouped together to form conceptually similar video units, called scenes. Our future work includes ontology support, incorporation of key-frame selection methodologies and the automatic creation of a list of semantic labels which will be proposed to the user, for the annotation of key-frames and shots, based on machine learning processes.

Acknowledgements. This work was undertaken in the framework of The History of Commandaria: Digital Journeys Back to Time project funded by the Cyprus Research Promotion Foundation (CRPF) under the contract ANTHRO/0308(BE)/04.

References

1. Papadopoulos, K., Tsapatsoulis, N., Lanitis, A., Kounoudes, A.: The History Of Commandaria: Digital Journeys Back To Time. In: 14th International Conference on Virtual Systems and Multimedia (VSMM 2008), Limassol, Cyprus (2008)
2. Benjamins, V.R., Contreras, J., Blázquez, M., Niño, M., García, A., Navas, E., Rodríguez, J., Wert, C., Millán, R., Dodero, J.M.: ONTO-H: A collaborative semiautomatic annotation tool. In: 8th International Protégé Conference Collaborative Development of Ontologies and Applications, Madrid, Spain, pp. 69–72 (2005)
3. Lin, C.Y., Tseng, L., Smith, R.: Video Collaboration Annotation Forum: Establishing Ground-Truth Labels on Large Multimedia Datasets. In: NIST Text Retrieval Conference, TREC (2003)
4. Schreiber, A.T., Dubbeldam, B., Wielemaker, J., Wielinga, B.: Ontology-Based Photo Annotation. IEEE Intelligent Systems 16(3), 66–74 (2001)
5. Fujisawa, S.: Automatic Creation and Enhancement of Metadata for Cultural Heritage. Bull. IEEE Tech. Committee on Digital Libraries (TCDL) 3(3) (2007)
6. ISO/IEC 15938-3:2001 Information Technology - Multimedia Content Description Interface - Part 3: Visual, Version 1
7. ISO/IEC 15938-4:2001 Information Technology - Multimedia Content Description Interface - Part 4: Audio, Version 1
8. ISO/IEC 15938-5:2003 Information Technology - Multimedia Content Description Interface - Part 5: Multimedia Description Schemes, First Edition
9. Ricoh Movie Tool, website http://www.ricoh.co.jp/src/multimedia/MovieTool
10. Adams, W.H., Lin, C.Y., Iyengar, B., Tseng, B.L., Smith, J.R.: IBM Multimedia Annotation Tool. Technical report, IBM Alphaworks (2002)
11. Bargeron, D., Gupta, A., Grudin, J., Sanocki, E.: Annotations for Streaming Video on the Web: System Design and usage Studies. In: ACM 8th Conference on World Wide Web, pp. 1139–1153. Elsevier, North-Holland (1999)
12. Steves, M.P., Ranganathan, M., Morse, E.L.: SMAT: Synchronous Multimedia and Annotation Tool. In: 34th Hawaii International Conference on Systems Sciences (2001)
13. Nack, F., Putz, W.: Semi-automated Annotation of Audio-Visual Media in News. GMD Report 121 (2000)
14. Cultural Heritage Online (ECHO), http://www.mpi.nl/echo/
15. Lienhart, R.: Comparison of Automatic Shot Boundary Detection Algorithms. In: SPIE Storage and Retrieval for Image and Video Databases VII, San Jose, CA, USA, vol. 3656, pp. 290–301 (1999)
16. Goldberg, D.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading (1989)