Switching Partners: Dancing with the Ontological Engineers

Werner Ceusters

Professor of Psychiatry, SUNY Buffalo, and

Director, Ontology Research Group, New York State Center of Excellence in Bioinformatics and Life Science.

Barry Smith

Julian Park Distinguished Professor of Philosophy, SUNY Buffalo

> A certain measured, cadenced step, commonly called a "dancing step", which keeps time with, and as it were beats the measure of, the Music which accompanies and directs it, is the essential characteristic which distinguishes a dance from every other sort of motion. (A. Smith 1980, 207)

Introduction

"Ontology" is a term increasingly used in all areas of computer and information science to denote, roughly, a hierarchically organized classification system associated with a controlled, structured vocabulary that is designed to serve the retrieval and integration of data. An ontology under this view is an artifact whose purpose is to ensure that information about entities in some domain is communicated successfully from one context to another, and this despite differences in opinions about what is the case in that domain or differences in the terminology used by the authors to describe the entities it contains.

Ontologies are today being applied in almost every field where research and administration depend upon the alignment of data of distributed provenance. They are being used, for example, by biologists to classify genes, toxins and proteins, and by medical scientists to classify diseases, drugs, therapies, and body parts. An example of the latter is the Foundational Model of Anatomy (FMA) which is an ontology of normal adult human and mammalian anatomy. Figure 1 shows the classification of "left leg" in the FMA in which it is categorized, among other things, as being a body part which is part of the left lower limb of either a male or female body.

Ontologies are making inroads also in the wider culture. There is an explosion of so-called "folksonomies" used to tag images on the web. The CIDOC ontology is being used by museum authorities to classify cultural artefacts (Doerr 2003). Ontologies have also been developed to assist lawyers in resolving disputes over the nature of patent and copyright and in determining how different versions of musical or literary works are to be treated for purposes of intellectual property protection (Ceusters and Smith 2007).

Ontologies of the kind just sketched are primarily used directly by humans to perform some classification task, as for example to provide appropriate general descriptors for organizing scientific papers in a library collection. Users of the library, on the other hand, can use this same ontology to find papers on topics they are interested in. An example of this sort of ontology, for web pages rather than scientific papers, is the DMOZ ontology from the Open Directory Project (*Open Directory Project*). However, ontologies are increasingly being designed to support computer-based services directly --

which means: without human intervention. Among the earliest such applications were ontologies applied to text-mining tasks such as automated indexing, topic extraction and summarization of information presented in textual format. Here we will focus on another illustration of the way in which ontologies are being used to help unlock the secrets of human culture, an illustration drawn from the domain of human bodily movements.

<space>

The ontology of motion

Our movements are being captured on video, and considerable resources are being invested in the development of techniques to extract information about such movements from the digital outputs of video surveillance cameras. Human movements can be classified, in the first place, from the purely kinematic point of view. But what are kinematically the same movements may still need to be classified in entirely different ways because they occur in different contexts. Consider a short movement of one lower leg crossing the other leg with the foot pointing outwards. Such a movement can be part of a mannequin’s step on the cat-walk, an epileptic jerk of the lower leg, the kicking of a ball by a soccer player, a signal ("get out!") issued in heated conversation, or a "half cut" in Irish *Sean-nós* dancing (Figure 2). When we focus on dance movements, image classification is made all the more difficult by the fact that, while we are dealing here with kinematic phenomena which are constrained, in complex ways, by systems of rules, these rules are themselves artefacts of culture which are marked by complex spatial and temporal variations. Like the cultural artefacts which they govern, they are subject to a continuous process of evolution. A host of additional problems are created for software agents designed to "understand" what is displayed in video images, by the need to translate information about the changes in the pixel configurations that constitute such images into information about movements carried out by the corresponding entities in reality. Can ontological engineering help us to understand what dancing is all about? As a first step towards answering this question we shall explore some of the problems faced when we seek to create software which can help the machine to recognize what dance is being performed in a particular video.

<space>

Pushing the boundaries of information retrieval

Dance has served throughout history as an important force for social cohesion, and public interest in social dancing is once again booming. Western Europe, in particular, is witnessing a huge revival of interest in folk dancing, which means: dancing of a sort which is rooted in the culture of some population but has undergone certain characteristic families of changes in the course of time. These dances include especially those which in countries such as France, Belgium, Germany and The Netherlands are called *bal folk dances* -- thereby referring to the fact that they are danced at events which are called a "*bal folk*". They comprise dances such as polka, mazurka, Scottish, *an dro*, *hanter dro*, *bourrée, branle*, (old time) waltz, *chapelloise* (Aleman’s *marsj*), *cercle circassien*, which are becoming increasingly popular amongst dancers of all age groups, complementing a no less intensively burgeoning interest in modern "ballroom dancing," which comprises dance types such as the waltz, foxtrot, American tango, and so forth. Social dancing events such as *Le Grand Bal de l’Europe* in France (*Le Grand Bal de l'Europe*), *Andanças* in Portugal (*Andanças*), *Gran Bal Trad* in Italy (*Gran Bal Trad*) and the Dance Flurry in Saratoga Springs (Dance Flurry Organization 2007), to mention just the most important ones, each attract several thousand dancers every year.

There is clearly a need, supported by a broad and still growing community, to gain a better insight into this complex of dance cultures and associated traditions and also to make sure that it is preserved in its full richness for the future. Growing in tandem with this need is an enormous demand for more and better information about such dances, not only from individual dancers, dance organisations and dance historians, but also from cultural agencies, libraries, tourist organizations, and dance teachers. Questions relate to the origins of these dances, to the rules governing how to dance them, to the different types of variations across space and time, and thus also to the associated question of when a given dance can properly be referred to as "the same" as a dance popular some centuries earlier.

In answering such questions, traditional dance resources and archives fall short -- and this is so even where the relevant information is made available online. If one is not an expert in dance history or choreography, which most social dancers are not, then it is nearly impossible to formulate a search question in such a way that the answers retrieved are relevant and useful to the individual dancer. Search engines that include video resources in their search space may for example perform well when it comes to retrieving videos in which a specific dance such as a Viennese Waltz is displayed -- but this is so only when the user includes the exact term "Viennese Waltz" in his query. This, however, is to suppose that the user already has command of the terminology relevant to the dance he is interested in and in the relevant language, something that is not always the case.

<space>

Search scenarios

Google can readily provide images in response to search inputs such as "Werner Ceusters" or "Barry Smith". But the much harder, and more interesting, challenge would be for a search engine to allow a user to submit an image and have the system inform him that the image is one of Werner Ceusters or of Barry Smith. Can we, in similar vein, envisage a search engine able to analyze fragments of video as a basis for providing information back to the user telling him what the video depicts? Here are a few scenarios: a tourist for example might bring home video fragments of people dancing in a *fest-noz* in Brittany and would like to know the names of the particular dances captured in the video so that he can go on from there to answer questions concerning the region of origin or choreography of the dance. Or a researcher, to take another example, is studying the evolution of specific dances over time and wishes to retrieve videos of dances with similar choreographies. More ambitiously, a historian might wish to assemble an entire evolutionary history of dances of a given type, with genealogical trees indicating influences. Or an American country dance choreographer might want to be sure that a commissioned creation will be sufficiently different from what already exist in the genre in question. She, for sure, would find it most helpful if she could submit a video of this creation to an intelligent video library that is able to find fragments of the included choreography in other dances. And then here too we address a characteristically ontological question: when do given dance fragments manifest the *same* choreography?

To exhibit useful sorts of behaviour in response to such challenges the system would need to bridge the semantic gap between the information that the computer can extract from given multimedia material and the interpretation that would be useful to a human user in a given pre-defined situation (Smeulders et al. 2000; Roach et al. 2002). This gap is very large. Imagine a video fragment which shows a person walking in the direction of the camera. Not only does the analysis software need to be able to identify the moving object as a person, it should also be able to avoid misinterpreting the enlargement of the group of pixels which depicts a given person as the person moves towards the camera to be the result of the person's growing larger. Bridging this gap seems to be achievable, but thus far only in certain very specific application domains, as for example is described in (Izquierdo et al. 2004), which describes a system that allows users to search for similar fragments in videos about soccer and tennis games. The similarity of such fragments is computed automatically on the basis of visual features such as colours and textures (Xu et al. 2004), a technique also used in the system described in (Christmas et al. 2001), where it was combined with search facilities based on auditory clues. (When a goal is scored in a soccer game, then at least one part of the audience leaves no doubt about their appreciation of that event by shouting and applauding.) Similar achievements have been reported in relation to tennis videos (Dahyot et al. 2003), and also for video news broadcasts (de Jong, Ordelman, and Huijbregts 2006). But, however impressive these results may seem, the problems still to be solved to be in a position to apply these same techniques to the domain of dancing remain enormous.

<space>

Setting the research agenda

These problems fall into at least two categories: one group involves the difficulties in representing the domain of dancing in a way that can be understood by software agents; the second involves problems in how to apply such representations to the semantic analysis specifically of that sort of content which is provided by video images. The latter is primarily an engineering issue; the former, however, requires the ontological engineer to crack a hard historicosociocultural nut: in order to "*represent*" dancing in the computer, we must first have a good insight into what dancing *is*. Before answering this question, we consider, first, the more narrowly engineering-related problems involved in the software analysis of video content.

<space>

Challenges in automatic video understanding

Automatic video-understanding is a relatively new field for which the research agenda has been set only fairly recently. (Cetin 2005) identified two so called "*grand challenges"* for video-analysis: the first was to develop applications that allow a natural high-level interaction with multimedia databases; the second was finding adequate algorithms for detecting and interpreting humans and human behaviour in videos containing also audio and text information. In addition, a number of intermediate-scale challenges relevant to successful human behaviour analysis have been identified. Some of these, such as problems inherent to face detection, might at first seem to be irrelevant in the context of dancing. They turn out, however, to be of crucial importance at least to dances of certain sorts. Facial expressions are in some cultures an intrinsic part of a dance; Argentine Tango dancers tend to look rather serious while moving over the dance floor, while ballroom champions invariably are smiling, though often in a way that is somewhat strained. Facial expressions might thus give additional clues about the sort of dance on display.

Automatic recognition of the type of dance displayed on a video requires facilities for detecting human bodies, their poses and postures, and their activities, and this even in spite of manifold variations in background (including for example differences between indoor and outdoor backgrounds). It requires robust techniques to discriminate and track body parts (arms and legs) belonging to one particular body from those belonging to another body, despite the fact that in many cases dancers wear similar costumes. In fact, recognizing objects in images and video is still an unresolved open problem in general and one of the main topics of research in content-based image retrieval (CBIR).

With respect to storage and retrieval, multimedia databases with semi-automatic or automatic natural interaction features do not exist. In this area, the ACM Multimedia Special Interest Group (SIG) which was created over ten years ago, recently identified two further grand challenges relevant in the context of analysing dancing videos (Rowe and Jain 2005). The first is to make the authoring of complex multimedia publications as easy as using a word processor or drawing program. Certainly there are a number of individual high-quality software packages for specific sub-tasks, but there is a conspicuous lack of seamless integration. The second challenge is that of bringing about a situation in which capturing, storing, finding, and using digital media will be everyday occurrences in our computing environment. The example provided "*show me the shot in which Jay ordered Lexi to get the ball*", is of the same nature as the kind of services that should be offered with respect to dance. This requires techniques that push paradigms such as motion-based classification and segmentation much further than currently realised (IEEE Computer Society 2005). Because digital representations of bodily movements are nothing more than groups of pixels appearing and disappearing in sequence, recognizing such movements in isolation, or recognizing movement-sound complexes such as *tap*, *stomp*, *clap*, *scuff*, and so on, requires insight into how such phenomena are captured in digital representations such as videos. Further aspects relevant for video analysis involve the ability to apprehend, classify and track backgrounds, costumes, numbers of couples dancing, camera positions, and so forth. Feature extraction algorithms are as yet insufficiently mature to capture subtle differences between movements such as a "ground cut" and a "half cut" in Irish *Sean-nós* dancing, or between a *polska* and a *hambo* turn in Swedish traditional dances. This is not only a matter of pixel granularity, which may be insufficient to allow capture of the necessary detail (such as foot movements of a particular couple dancing in a crowd), but also a problem of knowing what is there to be captured in order that conclusions of the required sort can be drawn.

A system with the capacity just sketched should, ultimately, be able to recognise not only what dance is being performed but also whether it is experts or beginners who are dancing. It should be able to identify the various phases in the dance (such as the specific figures in a Scottish *ceilidh*, or the successive moves in a contra dance, and even very short meaningful fragments such as the execution of a *pas de bourrée* in ballet). It should also be able to detect differences in style, such as between the old and the modern Sevillana dancing styles (both still danced socially today) in contrast to, for instance, *bourrées* where the older form, dating back some hundred years, would now only occur in dance performances rather than in social dancing. In addition, *bourrées* are danced differently in different parts of Europe. Clearly, such differences have to be clarified before video analysis becomes possible at this level at all -- which brings us to the second, more properly ontological, challenge to be addressed, namely: what is dance, and how can dances be classified?

<space>

Challenges in understanding dancing

Dancing has been a subject of research for thousands of years and from a variety of different perspectives. The questions which interest us here concern: (1) what is dancing and how can it be distinguished from other complex forms of human behaviour (or combined therewith as in the martial art *capoeira*) and (2) how have specific dance types evolved culturally over time?

UNESCO classifies dance as belonging to the what it calls "intangible heritage", which consists of creations originating in a given community and based on oral traditions, customs, languages, music, rituals, festivities, traditional medicine and pharmacopoeia, the culinary arts and all kinds of special *skills* connected with material aspects of culture, such as those involving tools and the habitat (UNESCO 2007). This definition is part of the Convention for the Safeguarding of the Intangible Cultural Heritage adopted by the 32nd session of the UNESCO General Conference in 2003 (UNESCO 2003), where the term "intangible heritage" supersedes the older term "folklore". The earlier folklore model supported scholars and institutions in documenting and preserving a record of disappearing traditions. The more recent model of intangible heritage aims rather to sustain a living tradition by supporting the conditions necessary for cultural reproduction. This means according value to the “carriers” and “transmitters” of traditions, as well as to their habitus and habitat. The task is to sustain the whole system as a living entity and not just to collect “intangible artifacts.”

A further dimension which enters here is that of socioeconomic factors having to do with the different ways in which dance, like other forms of culture, interacts with the wider social order. In the cultural experience of Europe, for example, and as contrasted with the North American case, cultural forms were usually generated by and aimed at cultural elites, becoming transmitted to larger swathes of the population only progressively and after some elapse of time. This elitist pattern constitutes the backbone of the European cultural tradition in domains as various as religion, music, eating habits, dress, manners, and daily life in general (Liehm 2002). And it holds too for dancing which, while being initially a focus of display among elite society, still rapidly came to enjoy appreciation across all levels of the social hierarchy. As we can learn from the history of social life and culture in Glasgow for example, Scottish country dancing began as an elite activity whose refinements were overseen by the various dance academies; the latter were however unable to keep pace with the “penny reel” gatherings that were becoming increasingly popular at fair-time. By the 1830s local householders were cashing in on the crowds attending Glasgow Fair by making their homes available for dances at a penny a time. Dancing was on its way to becoming a marketable commodity with mass appeal (King 1987). But as dances moved further from their cultural roots, there were inevitable clashes between traditionalists and innovators. As Trenner puts it: "Old-timers are motivated by their loyalty to the history of, the techniques of, and subtle sophistication of their forms. Newcomers are propelled by their enthusiasm, and will provide structure even when they have little information to guide them" (Trenner 1998).

Dance, like culture in general, is always changing, and it has to change in order to remain meaningful from one generation to the next. As current historiography teaches us, our past and our heritage are not things preserved for all eternity but processes that must constantly revalidate themselves. The successful, living aspects of culture produce new experiences for its users. Change in the way dances are performed is a matter not just of the passing of time but also of a progressive delocalisation -- they are both a part of the identity of each separate region but also incrementally evolving ingredients in a universal artistic language. They contribute on the one hand to the blossoming of cultural diversity and to the enrichment of specific cultural identities, while on the other hand their plasticity renders them capable of nourishing the dialogue between and intermingling of the various cultures (Rouger and Dutertre 1996). Social dancing therefore forms one of the cultural areas that is best adapted to achieving the cultural objectives set out in the "Treaty on European Union, Article 151 (ex 128)", which include: "to contribute to the flowering of the cultures of the Member States, while respecting their national and regional diversity and at the same time bringing the common cultural heritage to the fore" and "encouraging cooperation between Member States and, if necessary, supporting and supplementing their action in the following areas: improvement of the knowledge and dissemination of the culture and history of the European peoples; conservation and safeguarding of cultural heritage of European significance; non-commercial cultural exchanges; artistic and literary creation, including in the audiovisual sector" (European Commission 2006).

<space>

"Ontologies" and "ontology"

Earlier we presented a view of ontologies as representational artifacts that, when designed in appropriate ways, can help humans and software agents in performing classification tasks. The question here is whether they can also be used in the context of analyzing the content of dancing videos by bridging the semantic gap between pixel sequences of classificatory content. We will now present some reasons to believe that this is so. In (Hakeem and Shah 2004) it was shown that for the analysis of videos on which company board and project meetings are displayed, the most promising way to bridge the semantic gap is by using an ontology. To support analysis of video content in the domain of dance, we shall need to create two ontologies, the first describing real-world phenomena relevant for the domain of dancing itself, and the second covering how these phenomena are exhibited in videos through image and sounds. The former would require the generic parts of the ontology to cover relevant aspects of human motion, with more specific parts focused on the domain of dance motions in general, with more detailed components relating to the sorts of dances that are contained in the collection from which information is to be retrieved, including temporal indexing to enable capture of historical aspects of the ways these dances have evolved over time. Building ontologies of this level of detail and complexity poses a challenge in its own right as witnessed by the numerous examples of mistakes committed in the past (Ceusters and Smith 2003; Ceusters, Smith, and Goldberg 2005; Ceusters et al. 2004, 2004a). It is here that "ontology" as a scientific discipline, rather than a computational artifact, comes into its own.

"Ontology" is of course a term having its roots in philosophy, where it means, roughly, the science of being. For a long time ontologists working on information systems ignored the fruits of ontological research in philosophy, and thus they often recommitted errors of a characteristically philosophical sort, above all by confusing the classification of entities in reality with the classification of the words or data describing such entities. Increasingly, however, it is being recognized in at least certain circles of ontological engineering that data integration of a useful sort cannot be achieved merely by classifying the words or concepts which different groups of experts associate with different types of data. The problems created by the differences in word usage among such groups are indeed precisely what need to be solved with the aid of ontologies.

In addition, ontology builders often fail to take into account the fine details that are needed in order to make their representations conform to what is the case in reality, or they resort to representational languages or systems that are insufficiently expressive to capture such details. An example is the classification of "dance" in the Open Directory Project mentioned above, and of which a small portion is shown in Figure 3. There are several questions that might be raised here:

- is a waltz just a ballroom dance ? Apparently not, as it is found in many of what ODP calls “folk and traditional dances” too.
- can all dancing activities really be considered to be performances, let alone works of (performing) art?
- why distinguish between *modern* and *contemporary* dance if all the subtypes of the former are also all and only the subtypes of the latter?
- why are “waltz” and “flamenco” classified twice -- both directly and indirectly under “dance”?
- can only capoeira be danced for recreational purposes ?
- how can a particular event be at the same time an instance of both a dance and a martial art ?

More fundamental questions, not clearly addressed in the ODP documentation are: what precisely is being classified here? Actual instances of dancing (dateable performances) or general types? As danced within a single culture or community, or across all of human culture? In a single era or across all of human history? To avoid building ontologies that lead to questions of this sort, a mechanism for making clear what the terms in the ontology are actually about, which means: what they represent on the side of reality, is imperative, as also is a clear account of the relation between the ontology and what it describes.

<space>

Ontology and dance

With respect to dance ontology, relevant work has been done by philosophers such as Roman Ingarden, whose *Ontology of the Work of Art* treats in succession the ontology of musical works, pictorial images, architecture and film on the basis of a general ontological theory of the structure of the work of art which distinguishes strictly between the work itself as a complex stratified object that is neither physical nor mental, and its various realizations in readings, performances, or physical artifacts (Ingarden 1989). For present purposes it is Ingarden's treatment of the work of music that is of most direct relevance. Here we can distinguish between the work itself, the score, the various performances, and the concretizations of the work in the experiences of listeners. For Ingarden it is important that the work itself, even though it is a bearer of identity from one realization to the next (and a benchmark for the faithfulness or adequacy of such realizations), nonetheless has a history (or what Ingarden calls a "life" -- in virtue of changes for example in performance style, instrumentation technology, and interpretations.

Of more direct relevance is the work of movement theorists such as François Delsarte, Frederick Matthias Alexander, Émile Jaques Dalcroze and Rudolph von Laban, all of whom developed influential techniques for thinking about movement. Delsarte was interested in enhancing dance pose and gesture through an understanding of the natural laws governing bodily movement. To that end, he carefully studied aspects of human gesture in everyday life and compiled records of thousands of gestures, each identified with specific descriptions of their time, motion, space and meaning (Shawn 1974). Alexander was an actor responsible for the educational process that is today called the Alexander Technique, a method of helping people learn to free up their habitual motor reactions through improvement of kinesthetic judgment (Alexander 1989). Von Laban was an Austrian-born architect, philosopher, and choreographer who developed a sophisticated system of movement observation and description (Laban Movement Analysis or LMA) that enables the observer to identify and articulate what parts of the body move, and when, where, and how they move. The body's relationship to "space," "shape," and "effort" (or inner impulses) are some of its primary elements; but in contrast to other methods LMA places no emphasis on which movement quality or shape is desirable from aesthetic or other perspectives (Hutchinson 1991).

Such movement analysis and annotation methods can be used to write out dances, an activity that is called *choreology*, methodologies for which were developed not only by von Laban but also by Feuillet, Stepanov (Nijinsky), and Benesh. As an example, Benesh developed a purely kinetic language which allows positions, steps, and other movements to be directly represented, including movements of multiple dancers involved in complex dance productions. Not only can the reader see the movements that are written in a score, the mode of recording is sufficiently realistic that they can also be felt motorically. The advantage of using a purely kinetic method for describing dance and movement is that it is the movement itself that is conveyed, rather than some analytical, functional, scientific or poetic verbal description (Benesh and Benesh 1983).

From the scientific perspective, folk or social dancing has generally been considered to be an incidental part of musicology and thus it has attracted little attention from scholars, who have seen it as lacking the prestige enjoyed by other aspects of musical life. Thus far, research in dance history has been primarily focused on dance culture during single epochs, for example in relation to important political events such as the Congress of Vienna of 1815. This has made it very difficult to have a clear understanding of the continuity or discontinuity in dance culture from one epoch or culture to another. It is indeed much easier to compile statistics about the numbers of dancers active at given times and in limited geographical areas than to gauge differences and similarities between, say, waltz choreography today and at different times in the past. Even state of the art research such as has been published by Monika Fink (Fink 1996), Richard Semmens (Semmens 2004) or Jean-Michel Guilcher (Guilcher 2004) fall far short of conveying an adequate picture of how dance nomenclature and terminology have evolved and whether or not the changes have adequately reflected simultaneous evolution of the dance types being studied. In addition, most projects have focused on individual dance practitioners as source material, and are thus biased. And because the results have been published mostly in the form of printed collections such as Valerie Preston-

Dunlop's *Dance Words* (Preston-Dunlop 1995), they cannot be searched efficiently, nor are they very useful for the analysis of dance practice, as they tend to be colored by historically and culturally specific theories of what dance *should* be, rather than what it actually *is* or *was* at specific stages in the past.

There is clearly room here for a more objective analysis which can be tested against the really existing needs of large communities of interested persons and in a way that will not only contribute to the quality and quantity of information available on-line but also yield deeper scientific insights for example as concerns global patterns in transmission of cultural phenotypes from one generation to the next.

<space>

Ontologies for video analysis, indexing and retrieval

An important component of the tools for video image understanding of the future will be the ability to make decisions based on a progressively closer approximation to a correct analysis of complex motions on the basis of successively more refined hypotheses as to the activities involved. Recent work has investigated the application of Hidden Markov Models (HMMs) in a layered approach to the recognition of individual and group actions on the basis of multi-modal recordings (McCowan et al. 2005). In this approach, a first layer integrates modalities and recognizes low-level elements such as motion patterns and tempo. A second layer takes likelihoods from the lower layer as input features, integrates them with features coming from audio-analysis, and generates hypotheses as to the nature of the actions involved, for instance as concerns the type of dance performed by the entire group and the level of expertise of the dancers. Initially, simplifying assumptions are made, which then need to be relaxed in order to address the complexities of real situations. In particular, combinations of activities may change over time, sometimes gradually, sometimes abruptly. The system must recognize and adapt to these changes. Predictive platforms have to learn to respond to such changes efficiently. Also, prediction algorithms need to be able to infer actions in the presence of multiple people engaging in the multiple sorts of complex actions that are involved in any given instance of social dancing.

Achieving efficiencies of this sort will involve the use of algorithms which can evolve in light of lessons learned in successive applications. Examples of such approaches include: applying high-level learning algorithms like neural networks (Gurney 2002), reinforcement learning (Mozer 2005), semi-supervised learning of Bayesian Network classifiers, and case-based reasoning in complex environmental settings (Cohen et al. 2004). Though each such approach can form the basis of the resolution of certain tasks that will be required in a powerful dance analysis system of the future, none as yet achieves high level recognition specifically targeting dancing events. Practical applications thus far are limited to areas such as video surveillance in railway stations (Cupillard et al. 2004), banks (Georis et al. 2004) or combat areas (Kalukin 2005), where the goal is to assess automatically coarse-grained phenomena such as crowding, blocking of entries, vandalism, and so forth. But these applications take advantage of the fact that video-surveillance cameras work with (more or less fixed) backgrounds and under conditions where it suffices to detect large-scale movements. The requirements for identifying what sorts of dances are being recorded on video are more demanding, though progress towards the necessary fine-grained analysis is being made. In (Kiranyaz et al. 2003), a system is described that is capable of indexing domainindependent large image databases and that allows retrieval via search and query techniques based on semantic and visual features. In the professional annotation module of the BUSMAN system, an advanced interface for automatic and manual image annotation has been developed, although the automatic annotation functionality is mainly based on low-level descriptors such as simple shapes, colours and textures (Waddington 2004). *MediaArchive*, from Blue Order (Blue Order 2007), is a powerful industrial archiving system allowing storage and retrieval of any media files and is currently used by several major broadcasters across Europe to give users access to their collections. However, much research is still required, the coordination of which is being attempted in Europe by the Network of Excellence in Content-Based Semantic Scene Analysis and Information Retrieval, SCHEMA (Kompatsiaris 2004).

Specifically ontological contributions to video annotation include (Nevatia, Hobbs, and Bolles 2004), which describes an event ontology framework, a formal representation language, and ontologies for the security and meeting domains. However, the representations of the domains selected are highly simplified and each of the developed ontologies contains not more than a dozen entities. Clearly, the framework needs to be refined in such a way as to be applicable to much more complex events and the ontologies need to be expanded significantly. Also extremely small in design is the TRECVID 2005 “ontology” (10 entities) (Over 2006). And the same applies to the LSCOM (Large Scale Concept Ontology for Multimedia), a work in progress on the part of IBM working together with Carnegie Mellon, Columbia University and the University of California at Santa Barbara, which envisages an ontology of the order of some 1000 entities, given that this ontology is designed to be used for “understanding” the entirety of news broadcast content (Smith, J. et al. 2005).

Moving in the direction of more structurally coherent ontologies, Bremond *et al*. have developed a video event ontology consisting of representations of entities of two main types: physical objects in an observed scene and states and events occurring in the scene (Bremond et al. 2004). The former are divided into *static objects* (such as a desk, a machine) and *mobile objects* detected by a vision routine (e.g. a person, a car). The latter correspondingly are divided into (static) primitive and composite *states*, and (dynamic) primitive and composite *events*. The authors use logical and spatial constraints to specify the physical objects involved in a scene and also temporal constraints, including Allen’s interval algebra operators, to describe relations e.g. of temporal order. The result was used as a framework for building two ontologies for visual monitoring of banks and of metro stations using ORION’s Scenario Description Language (Bremond et al. 2004).

A comprehensive review of existing content-based retrieval systems and video retrieval literature can be found in (Izquierdo 2003), and several such systems are currently available for use such as, for example, *Query By Image Content* (QBIC) (IBM Corporation 2007) which is an image retrieval system developed by IBM and currently in use at the Hermitage Museum for its online gallery (State Hermitage Museum 2003).

To make an endeavour of this magnitude succeed, which means: to make it possible to extract low-level (for example kinetic) features from videos in such a way that they can be combined into constructs of a higher order corresponding to bodily movements and gestures, requires that the different existing multimedia standards need to be bridged by ontology standards. Several recommendations issued by the World Wide

Web Consortium, above all the Resource Description Framework (RDF) and the Ontology Web Language (OWL), have a role to play from the ontological point of view in realizing this goal, but these languages are as yet insufficiently mature to be usable for the representation of complex spatial-temporal detail of the sort that is required in facing the challenges at issue here. They are also not yet easily combinable with the multimedia standards developed by the International Standards Organisation (ISO) and the International Electrotechnical Commission (IEC) such as JPEG 2000, MPEG-4, MPEG-7 and MPEG-21 when dealing with issues such as storing, transmission and editing of still and video images. Some of these standards are versatile enough to accommodate some of the multimedia access services as described above, but much work still needs to be done. Currently, audiovisual content described with MPEG-7 elements (description schemes and descriptors) are expressed solely in the language known as "XML Schema". The latter has been ideal for expressing the syntax, structural, cardinality and data typing constraints required by MEPG-7. It has also been used to build a preliminary dance ontology based on Labanotation (Hatol 2006). However, in order to make the descriptions accessible, re-usable and interoperable with other domains, the semantics of the MPEG-7 metadata terms need to be expressed in an ontology using a language like OWL. Therefore, new specific data types geared to multimedia content and media-specific data types have to be proposed. A first attempt to build a dance movement ontology called DVSM (for "Dance Video Semantics Model") is proposed in (Ramadoss and Rajkumar 2007). The goal of the authors is to use this ontology in the context of video-annotation, but it should be applicable also under some of the other scenarios described above.

<space>

Towards a semantic-web based infrastructure for experience-based search and retrieval for multimedia dance resources.

Delivery of and access to audiovisual content is a business that has accounted for a significant percentage of the world’s Gross Domestic Product in recent years and is growing continuously. Current stocks of audiovisual content are growing at an exponential rate and multimedia services are becoming increasingly sophisticated and heterogeneous. On the consumer side, video on demand, copying and redistribution, as well as retrieval and browsing through video portals as offered by Google, Yahoo and YouTube, is becoming progressively more popular. Although this development brings increased revenue for the telecommunications and content provider industry, it also brings a challenging task: to deliver customised functionalities for fast query, retrieval and access.

Building a digital repository and related analysis-, search- and retrieval tools for the domain of video representations of dance is indeed only a small fragment of the totality of what is required to meet this global challenge. Tackling this narrow fragment will however involve addressing many of the same technical problems as need to be addressed on the broader front, and will surely bring valuable lessons. At the same time it will also improve our knowledge of and access to an important aspect of our cultural heritage, and also lead to better research and education in this area. As an example: why are culturally educated people not at all shocked when in the 1954 version of the film *Brigadoon* English contradance choreographies are used to represent and evoke a Scottish dance event, or when a mazurka and non-Regency quadrille are danced in *Pride* *and Prejudice*? The latter is no less of an anachronism than would be involved in using Monet paintings or Rolex watches in a film about the life of Mozart. Nevertheless, the former would today pass unnoticed, while the latter would be considered an insult to our historical consciousness. Stimulating a systematic quest for data and analyses of collected data around dance culture will help to bring about a rectification of this kind of situation, with broader consequences for our level of sophistication about our own historical past.

The system would be a tremendous help for analysing both the form (the danced steps) and the ways in which this form becomes meaningful to its users (the meanings found for example in ethnic and national dances). As such, it will further contribute to the critical re-evaluation of dance, and through dance, of our wider cultural heritage. Dances can be illustrative of cultural trajectories, influences, fashions, and traditions of an entire continent. Understanding and learning the dances of an alien culture can help us to understand this culture in new ways. A system with the capacities that we described will also foster the development of new traditions by bringing new cultural influences that may be both temporally and spatially distant (i.e. looking at pictures of a waltz in the nineteenth century can give us new ideas as to how to dance it today, how it is related to other dances, and so on).

<space>

Conclusion

Developments designed to allow video search in areas such as dancing are crossing disciplinary boundaries particularly between the arts and humanities on the one hand and technology on the other. Innovative ideas can only derive from a tight integration of professionals from different fields, so that the appropriate state-of-the-art technology is not simply put to use by end users but actually designed with them and for them. As research and publication increasingly move to the Internet, so research materials, too, have to become more accessible via computerised interfaces, and research in the arts and humanities has to become more efficient. Software applications such as the ones envisaged here will enable speedier recovery of data and facilitate its analysis in ways which will assist both archiving of and research on dance. The ontologies that have to be produced as the basis for such software applications should be made openly accessible to researchers. Research questions raised in conjunction with such developments can be adapted to other fields, hopefully also with innovative results. Indeed, the vision algorithms and theories that have to be developed to make such searches possible can be re-used for other scientific and practical purposes in all contexts where a video corpus has to be analysed and queried.

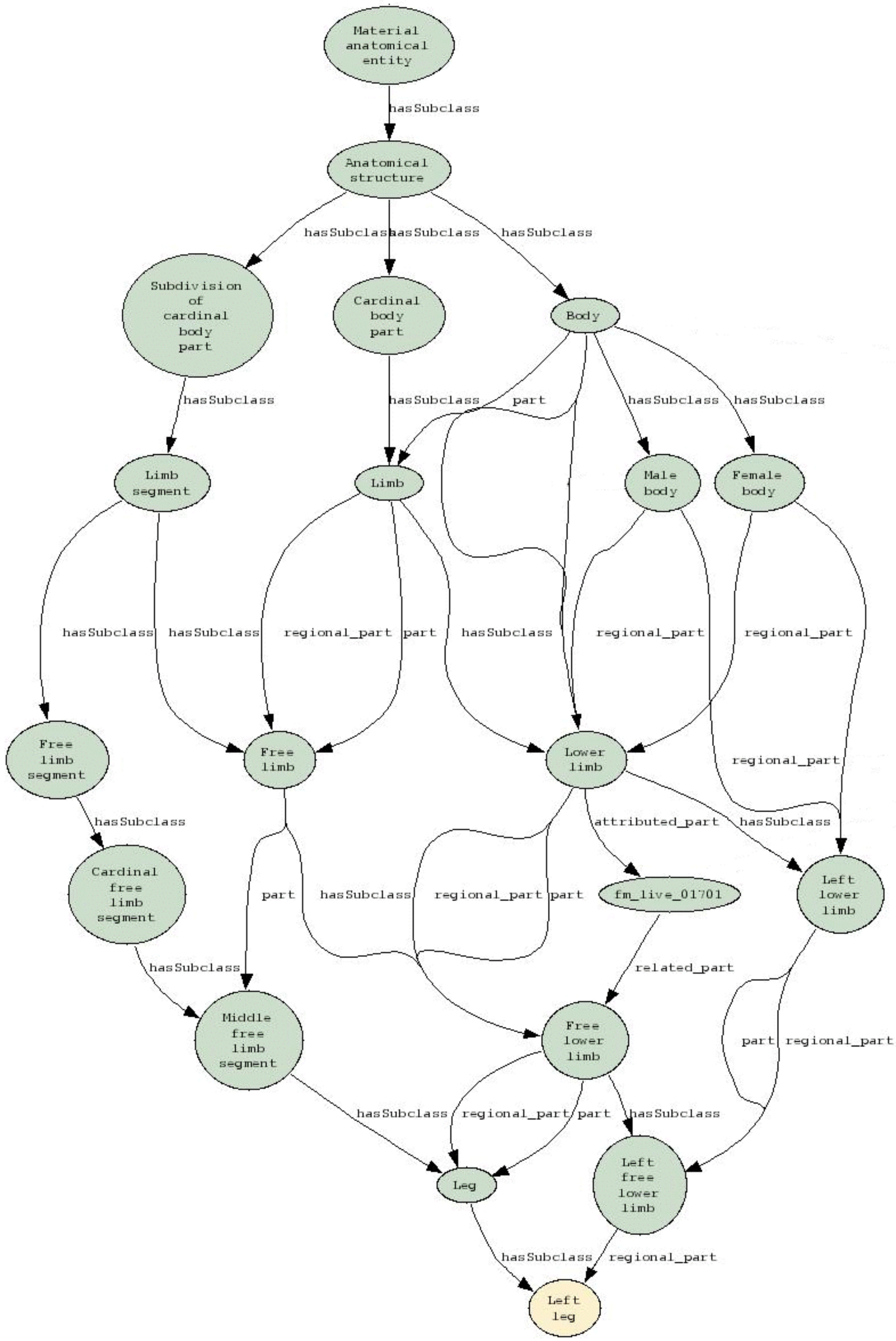
Material anatomical entity
hasSubclass
Anatomical structure
hasSubclass
hasSubclass
hasSubclass
Subdivision of cardinal body part
Cardinal body part
Body
hasSubclass
hasSubclass
part
hasSubclass
hasSubclass
Limb segment
Limb
Male body
Female body
hasSubclass
hasSubclass
regional_part
part
hasSubclass
regional_part
regional_part
Free limb segment
Free limb
Lower limb
regional_part
hasSubclass
attributed_part
hasSubclass
Cardinal free limb segment
part
hasSubclass
regional_part
part
fm_live_01701
Left lower limb
hasSubclass
related_part
Middle free limb segment
Free lower limb
part
regional_part
hasSubclass
regional_part
part
hasSubclass
Leg
Left free lower limb
hasSubclass
regional_part
Left leg

Figure 1: Classification of the entity "*Left Leg*" in the Foundational Model of Anatomy.

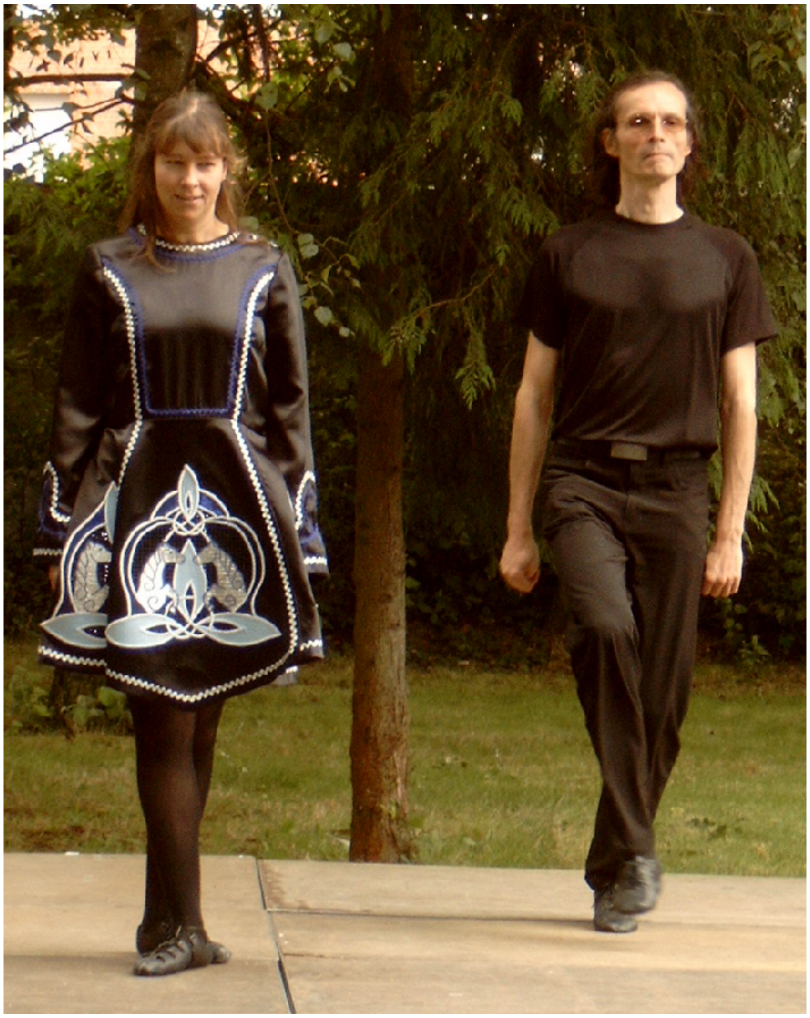

Figure 2: Irish *Sean-nós* dancer (right) doing a "half cut" with the right leg.

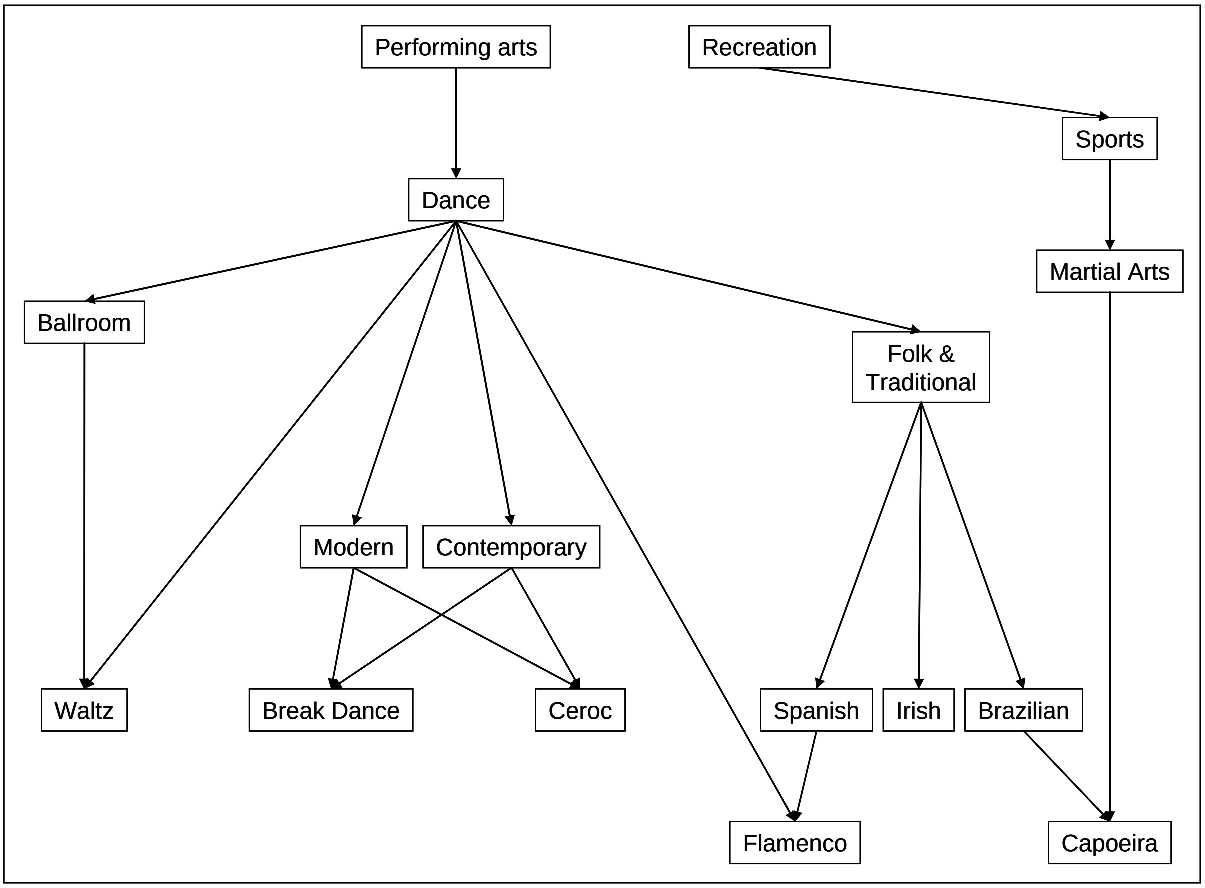

Figure 3: Part of the classification of "dance" in the Open Directory Project.

References

Alexander, F. Matthias, ed. 1989. *The Alexander Technique: The Essential Writings of F. Matthias Alexander*. Edited by E. Maisel. New York: Citadel Press.

Andanças 2007. [cited November 1 2007]. Available from http://www.pedexumbo.com/index.php?option=com_content&task=view&id=26&Itemid=59.

Benesh, Rudolf, and Joan Benesh. 1983. *Reading Dance: The Birth of Choreology*: McGraw-Hill Book Company Ltd.

Blue Order. 2007. *Media Archive* 2007 [cited November 1 2007]. Available from http://www.blue-order.com/products_media_archive_professional.html.

Bremond, François, Nicolas Maillot, Monique Thonnat, and Van-Thinh Vu. 2004. Ontologies For Video Events.

Cetin, E. 2005. Interim report on progress with respect to partial solutions, gaps in know-how and intermediate challenges of the NoE MUSCLE.

Ceusters, Werner, and Barry Smith. 2003. Ontology and Medical Terminology: why Descriptions Logics are not enough. In *Towards an Electronic Patient Record (TEPR 2003)*. San Antonio.

———. 2007. Referent Tracking for Digital Rights Management. *International Journal of Metadata, Semantics and Ontologies* 2 (1):45-53.

Ceusters, Werner, Barry Smith, and Louis Goldberg. 2005. A terminological and ontological analysis of the NCI Thesaurus. *Methods of Information in Medicine* 44:498-507.

Ceusters, Werner, Barry Smith, Anand Kumar, and Christoffel Dhaen. 2004. Mistakes in medical ontologies: Where do they come from and how can they be detected? In *Ontologies in Medicine. Studies in Health Technology and Informatics*, edited by D. M. Pisanelli. Amsterdam, The Netherlands: IOS Press.

———. 2004a. Ontology-based error detection in SNOMED-CT®. In *MEDINFO 2004*, edited by M. Fieschi, E. Coiera and Y.-C. J. Li. Amsterdam, The Netherlands: IOS Press.

Christmas, William J., Josef Kittler, Dimitri Koubaroulis, Barbara Levienaise-Obadia, and Kieron Messer. 2001. Generation of Semantic Cues for Sports Video Annotation. In *International Workshop on Information Retrieval*. Oulu, Finland.

Cohen, I, N Sebe, F.G Cozman, M.C Cirelo, and T.S Huang. 2004. Semi-supervised Learning of Classifiers: Theory and Algorithm for Bayesian Network Classifiers and Applications to Human-Computer Interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (12):1553-1567.

Cupillard, F., A. Avanzi, F. Brémond, and M. Thonnat. 2004. Video Understanding for Metro Surveillance. In *The IEEE ICNSC 2004 in the special session on Intelligent Transportation Systems*. Taiwan.

Dahyot, Rozenn, Anil Kokaram, Niall Rea, and Hugh Denman. 2003. Joint audio visual retrieval for tennis broadcasts. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*.

Dance Flurry Organization. 2007. *The Dance Flurry* 2007 [cited November 1 2007]. Available from http://www.danceflurry.org/.

de Jong, F.M.G, R.J.F Ordelman, and M.A.H Huijbregts. 2006. Automated speech and audio analysis for semantic access to multimedia. In *First International Conference on Semantic and Digital Media Technologies, SAMT 2006, Lecture Notes in Computer Science 4306*. Athens, Greece: Springer Verlag.

Doerr, Martin. 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI Magazine archive* 24 (3):75 - 92.

European Commission. 2006. Treaty on the European Union, article 151 (ex 128).

Fink, Monika. 1996. *Der Ball. Eine Kulturgeschichte des Gesellschaftstanzes im 18. und 19. Jahrhundert*. Innsbruck: Bibliotheca Musicologica.

Georis, B., M. Mazière, F. Brémond, and M. Thonnat. 2004. A Video Interpretation Platform Applied to Bank Agency Monitoring. In *The Intelligent Distributed Surveillance Systems Workshop*. London, UK.

Gran Bal Trad 2007. [cited November 1 2007]. Available from http://www.granbaltrad.it/en/indexen.html.

Guilcher, Jean-Michel. 2004. *La Contredanse : Un tournant dans l'histoire française de la danse*, *Territoires De La Danse*. Paris: Complexe.

Gurney, K. 2002. *An Introduction to Neural Networks*. New York: Routledge.

Hakeem, Asaad, and Mubarak Shah. 2004. Ontology and Taxonomy Collaborated Framework for Meeting Classification. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*

Hatol, J. 2006. MovementXML: A Representation of Semantics of Human Movement based on Labanotation, School of Interactive Arts and Technology, Simon Fraser University.

Hutchinson, Ann, ed. 1991. *Labanotation: The System of Analyzing and Recording Movement*. 3rd ed. New York: Routledge/Theatre Arts Books.

IBM Corporation. 2007. *IBM's Query By Image Content* 2007 [cited October 29 2007]. Available from http://wwwqbic.almaden.ibm.com/.

IEEE Computer Society. 2005. 7th IEEE Workshop on Applications of Computer Vision / IEEE Workshop on Motion and Video Computing, 5-7 January 2005, at Breckenridge, CO, USA.

Ingarden, Roman. 1989. *The Ontology of the Work of Art (translated by Raymond Meyer with Jon T. Goldthwait)*. Athens, Ohio: Ohio University Press.

Izquierdo, E. 2003. State of the art in content-based analysis, indexing and retrieval.

Izquierdo, Ebroul, Ivan Damnjanovic, Paulo Villegas, Li-Qun Xu, and Stephan Herrmann. 2004. Bringing user satisfaction to media access: the IST BUSMAN Project. In *Proceedings of the Information Visualisation, Eighth International Conference on (IV'04)*: IEEE Computer Society.

Kalukin, Andrew. 2005. Automating camera surveillance for social control and military domination. *Online Journal*, http://www.onlinejournal.org/Special_Reports/042905Kalukin/042905kalukin.html.

King, Elspeth. 1987. Popular culture in Glasgow. In *The Working Class in Glasgow, 1750-1914*, edited by R. A. Cage.

Kiranyaz, S., K. Caglar, E. Guldogan, O. Guldogan, and M. Gabbouj. 2003. MUVIS: A Content-based multimedia indexing and retrieval framework. In *Third*

International Workshop on Content-Based Multimedia Indexing, CBMI 2003. Rennes, France.

Kompatsiaris, I. 2004. The SCHEMA NoE Reference System. In *Workshop on "Novel Technologies for Digital Preservation, Information Processing and Acess to Cultural Heritage Collections*. Ormylia, Greece.

Le Grand Bal de l'Europe 2007. [cited October 30 2007]. Available from http://gennetines.org/.

Liehm, Anthony J. 2007. *The Cultural Exception: Why?* 2002 [cited October 29 2007]. Available from http://www.kinema.uwaterloo.ca/liehm962.htm.

McCowan, Iain, Daniel Gatica-Perez, Samy Bengio, Guillaume Lathoud, M Barnard, and Dong Zhang. 2005. Automatic Analysis of Multimodal Group Actions in Meetings. *Pattern Analysis and Machine Intelligence* 27 (3):305-317.

Mozer, M.C. 2005. Lessons from an Adaptive House. In *Smart environments: Technologies, Protocols, and Applications*, edited by D. C. R. Das: J. Wiley & Sons.

Nevatia, R., J. Hobbs, and B Bolles. 2004. An Ontology for Video Event Representation. In *Computer Vision and Pattern Recognition Workshop*.

Open Directory Project 2007. [cited October 30 2007]. Available from http://www.dmoz.org/.

Over, Paul. *Guidelines for the TRECVID 2005 Evaluation*, 24 Jan 2006 [cited. Available from http://www-nlpir.nist.gov/projects/tv2005/tv2005.html.

Preston-Dunlop, Valerie. 1995. *Dance Words*. Newark: Harwood Academic/Gordon & Breach.

Ramadoss, B, and K Rajkumar. 2007. Modeling And Annotating The Expressive Semantics Of Dance Videos. *International Journal of Information Technologies and Knowledge* 1:137-146.

Roach, M., J. Mason, L-Q. Xu, and Fred. W. M. Stentiford. 2002. Recent trends in video analysis: a taxonomy of video classification problems. In *6th IASTED International Conference on Internet and Multimedia Systems and Applications*. Hawaii.

Rouger, Jany, and Jean-François Dutertre. 1996. The traditional musics in Europe; the modernity of traditional music. In *Music, culture and society in Europe*, edited by P. Rutten: European Music Office.

Rowe, Lawrence A., and Ramesh Jain. 2005. ACM SIGMM retreat report on future directions in multimedia research. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)* 1 (1):3-13.

Semmens, Richard. 2004. *The "bals publics" at the Paris Opera in the eighteenth century*. Hilsdale, NY: Pendragon Press.

Shawn, Ted. 1974. *Every Little Movement: A Book About François Delsarte*. New York: Dance Horizons.

Smeulders, Arnold W.M., Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. 2000. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (12):1349-1380.

Smith, Adam. 1980. *Essays on philosophical subjects*. Oxford: Oxford University Press.

Smith, John R., Murray Campbell, Milind Naphade, Apostol Natsev, and Jelena Tesic. 2005. Learning and Classification of Semantic Concepts in Broadcast Video. In *International Conference on Intelligence Analysis*. McLean, VA.

State Hermitage Museum. 2007. *The State Hermitage Museum: digital collection* 2003 [cited November 1 2007]. Available from www.hermitagemuseum.org/fcgi-bin/db2www/qbicSearch.mac/qbic?selLang=English.

Trenner, Daniel. 2007. *Modern Social Tango: The Changing of the Codes* 1998 [cited November 1 2007]. Available from http://www.danieltrenner.com/daniel/ar_codes.html.

UNESCO. 2003. Text of the Convention for the Safeguarding of Intangible Cultural Heritage.

———. 2007. *Intangible Heritage* 2007 [cited November 1 2007]. Available from http://portal.unesco.org/culture/en/ev.php-URL_ID=2225&URL_DO=DO_TOPIC&URL_SECTION=201.html.

Waddington, Simon. 2004. The BUSMAN Project. *IEE Communications Engineer*, August/September, 40-43.

Xu, Li-Qun, Paulo Villegas, M. Diez, Ebroul Izquierdo, Stephan Herrmann, V. Bottreau, Ivan Damnjanovic, and D. Papworth. 2004. A user-centred system for end-to-end secure multimedia content delivery: from content annotation to consumer consumption. Paper read at Third International Conference, CIVR 2004, July, at Dublin, Ireland.