# A Study on Metadata Extraction, Retrieval and 3D Visualization Technologies for Multimedia Data and Its Application to e-Learning

Naofumi YOSHIDA*

In this paper we discuss on multimedia database technologies especially for a metadata extraction method for video streams, a retrieval method by keywords and a given strategies, and a data mining method by 3D visualization method and its application for e-learning. The main features of the methods are: (1) immediate and automatic extraction of metadata by giving semantics to combinations of heterogeneous sensors for video streams, (2) flexible retrieval according to the selected strategy from broadening, deepening and expanding, for metadata of video streams, (3) automatic 3D visualization for holistic and detail relationships of video streams. In this paper we discuss the methods on multimedia technologies and its feasibility by experimental results.

**Key words**: multimedia, database, metadata, retrieval, data mining, visualization

## 1. Introduction

Metadata extraction, retrieval and data mining are key technologies for realizing multimedia databases. It is very important to realize effective methods as these technologies for precise multimedia acquisition.

In the research area on multimedia databases metadata extraction (Sheth 1998)[9] (Westermann 2003)[13] is effective for utilization of video streams. Especially, it is effective to extract metadata from video streams of learning or meetings, because learning or meetings includes expert knowledge such as states of the arts in academic fields, discussion points for conclusions, or turning points of decisions. In this research area on metadata extraction for video streams, typical approaches are image processing (Brady 1982),[1] (Yilmaz 2006),[14] speech recognition (McTear 2002),[4] and semantics recognition (Seth 1998).[9] As a challenge of these metadata extraction of video streams, a metadata extraction method for meeting video streams (Jaimes 2004)[6] has been designed. Instead of metadata extraction by image processing with large processing time, if we can immediately extract metadata from video streams, we are able to extract expert knowledge from learning or meetings, especially we can find discussion points for conclusions or turning points of decision.

In this paper, we discuss an automatic and immediate metadata extraction method by heterogeneous sensors for video streams (Yoshida 2005).[18] The main feature of the method is immediate and automatic extraction of metadata by giving semantics to combinations of heterogeneous sensors for video streams on learning or meetings.

The 2nd key technology for multimedia databases is a retrieval method. We have a lot of retrieval method (Khoshafian 1995),[7] (Sheth 1998)[9] for multimedia data. Flexible retrieval methods enables us to increase chances for multimedia acquisition.

In this paper we review flexible retrieval method according to the selected strategy from broadening, deepening and expanding, for metadata of video streams (Shimizu 2003).[10]

The 3rd key issue on multimedia databases is data mining (Han 2000).[5] Data mining includes many analysis technologies such as visualization, hidden rule extraction, decision tree generation, time-series analysis. Visualization ((Lamping 1996),[2] (Mackinlay 1991),[3] (Robertson 1993),[8] (Uchihashi 1999)[12]) is the one of the important technology that generates the interactive presen-

* Assistant Professor, Faculty of Global Media Studies, Komazawa University

tation of digital images or movies for users to understand data.

In this paper we discuss an automatic generation method of visualization for both holistic and detail relationships of multimedia data (Yoshida 2003a),[15] (Yoshida 2003b),[16] (Yoshida 2004).[17] Three-dimensional (3D) visualization is effective for provision of relationships recognition of multimedia data because 3D space can be represent two-dimensional (2D) logical and temporal relationship at a glance. We also discuss the feasibility and effectiveness of visualizing holistic and detail relationships of learning materials by experiments in actual lectures.

## 2. An Automatic and Immediate Metadata Extraction Method

In this section, an overview of an automatic and immediate metadata extraction method by heterogeneous sensors for video streams (Yoshida 2005)[18] is shown. Figure 1 shows an overview of the method. We have four steps in the method.

- Step-1: Sensor Identification
  Detect sensor data from any sensors. At this time, each sensor and identified by ID manager (Fig. 1), Sensor data is recorded by Sensor Recorder (Fig. 1).
- Step-2: Sensor Data with Time Stamping
  Record sensor data to Sensor Database (Fig. 1). At this time, a time stamping process is performed for sensor data from sensors without time stamping.
- Step-3: Generation of Metadata
  In this step, sensor data with synchronization by time stamping and Sensor Semantics Database (given semantics for combinations of heterogeneous sensors, Fig. 1) is compared and metadata is output by Sensor Combination Manager in Fig. 1 when sensor data and any entries of the sensor semantics database are matched.
- Step-4: Storage for Metadata
  Sensor Recorder stores metadata generated in Step-3. Sensor Recorder stores both sensor data and metadata for any kinds of retrieval after metadata generation.

The process of metadata generation in Step-3 and generation of Sensor Semantics Database is important in the method. A hierarchy shown in Fig. 2 (Jaimes 2004)[6] for metadata extraction of video streams was already defined. By this layering, any sensor is included to the method, and

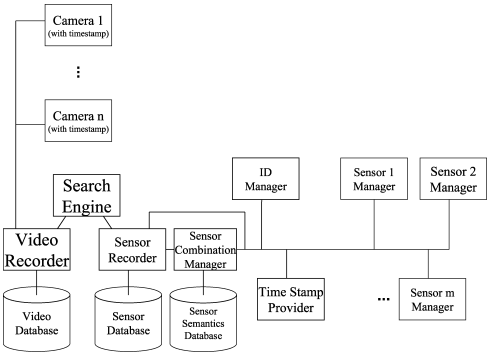


**Fig. 1.** An overview of an automatic and immediate metadata extraction method.

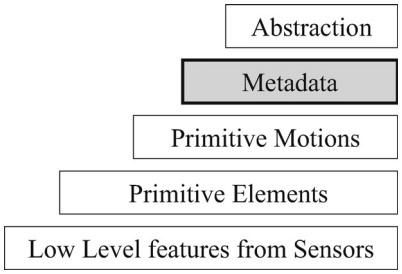


**Fig. 2.** Hierarchy for metadata extraction.

database designer of Sensor Semantics Database is manage semantics for sensors without detail knowledge of the sensors. Sensor Semantics Database is created independently of physical situation of sensors.

The first layer is for low level features from sensors. This layer represents raw data of sensors without time stamping.

Second, a layer for primitive elements above a layer for low level features from sensors is defined. This layer represents primitive elements with time stamping for specific applications.

A layer for primitive motions is defined for each logical units in time series of sensor data because video streams have transitions according to time and objective of the method is metadata extraction for video streams.

A layer for metadata is target of the method. In this layer, we are able to design metadata according to semantics of sensor data independently of layers for low level features from sensors, primitive elements.

A layer for abstraction is used by application of the method. Designers of applications of the

method will design them independently of sensor inputs, primitive elements, and primitive motions in video stream.

## 3. Retrieval Method According to the Selected Strategy from Broadening, Deepening and Expanding

In this section a retrieval method according to the selected strategy from broadening, deepening and expanding (Shimizu 2003),[11] for metadata of video streams is reviewed. Figure 3 shows an overview of this method.

### 3.1 Fomal expressions

Let $G$ is a set of the all concepts, $K$ is a set of the concepts that the user already knows, $N$ is a set of the concepts that that the user will acquire. $P_i$ are prerequisites of output data $i$, $Q_i$ are output concepts of data $i$. $r(x, Y)$ stands for a set of surrounding concepts of $x$ in $Y$, and $u(x, Y)$ stands for a set of lower concepts of $x$ in $Y$. Sets of concepts $S$ and $T$ is defined as follows:

$S = k \in N \mid \exists e \in K r(e, K)$

$T = k \in N \mid \exists e \in K u(e, K)$

Retrieval algorithm by strategy "Broadening" is defined as:

$(P_i \subset K) \wedge (Q_i \wedge N \neq 0)$

Retrieval algorithm by strategy "Deepening" is defined as:

$(P_i \subset K) \wedge (Q_i \wedge S \neq 0)$

Retrieval algorithm by strategy "Expanding" is defined as:

$(P_i \subset K) \wedge (Q_i \wedge N \neq 0) \wedge (Q_i \wedge T \neq 0)$

### 3.2 Implementation

Input of this retrieval method is three: ontologies that includes a set of concepts and their relations, degrees of understanding for users on each concetps, and retrieval candidates (multimeda data) with metadata of prerequisites and output concepts.

When a user submit queries, a user inputs keywords and selected strategy of from broadening, deepening and expanding. Then algorithm described in the section 3.1 is applied, then this method outputs orderd multimeda data as retrieval resutls.

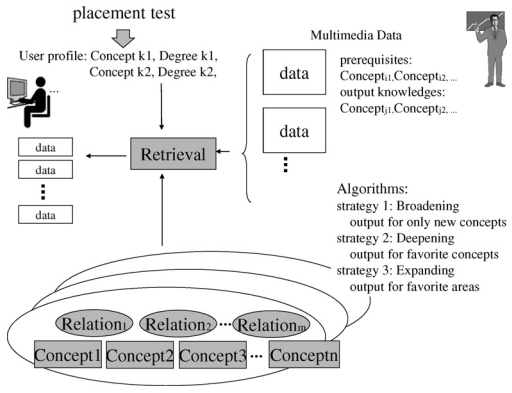


**Fig. 3.** An overview of a retrieval method according to the selected strategy from broadening, deepening and expanding.

## 4. Automatic Generation of 3D Visualization

### 4.1 Overview

An automatic generation method of 3D visualization (Yoshida 2003a),[15] (Yoshida 2003b),[16] (Yoshida 2004)[17] using XML Schema Definition (XSD) framework and schema compiler as shown Fig. 4.

The method enables to implement visualization method easily.

### 4.2 Implementation of automatic generation of visualization

The method enables to generate 3D visualization automatically for learners to recognize relationships of multimedia data.

The method is designed as following four steps:

- Step-1: Defining Schema by XSD (XML Schema Definition)
- Step-2: Compiling of Schema
- Step-3: Generation of XML Instance by Visualization Marshaller
- Step-4: Generation of Visualization Instance by Visualization Geometric Engine

### 4.3 Step-1: Defining schema by XSD

In this step, visualization designers will make definitions following two point of view by XML

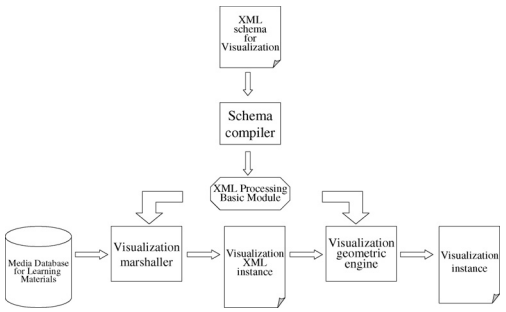


**Fig. 4.** An overview of system architecture for automatic generation of visualization.

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema" >
<xsd:element name="world">
                    <xsd:complexType>
                                        <xsd:sequence>
                                                            <xsd:element ref="candy" minOccurs="0" maxOccurs="unbounded"/>
                                        </xsd:sequence>
                    </xsd:complexType>
</xsd:element>

<xsd:element name="candy">
                    <xsd:complexType>
                                        <xsd:sequence>
                                                            <xsd:element ref="drop" minOccurs="0" maxOccurs="unbounded"/>
                                                            <xsd:element ref="tube" minOccurs="0" maxOccurs="unbounded"/>
                                                            <xsd:element ref="orientation"/>
                                                            <xsd:element ref="position"/>
                                        </xsd:sequence>
                    </xsd:complexType>
</xsd:element>

<xsd:element name="orientation">
                    <xsd:complexType>
                                        <xsd:sequence>
                                                            <xsd:element ref="position"/>
                                                            <xsd:element ref="radius"/>
                                        </xsd:sequence>
                    </xsd:complexType>
</xsd:element>

<xsd:element name="radius" type="xsd:float"/>

<xsd:element name="drop">
                    <xsd:complexType>
                                        <xsd:sequence>
                                                            <xsd:element ref="grain" minOccurs="0" maxOccurs="unbounded"/>
                                                            <xsd:element ref="arrow" minOccurs="0" maxOccurs="unbounded"/>
                                                            <xsd:element ref="relation-between-grain-and-arrow" minOccurs="0"/>
                                                            <xsd:element ref="position" minOccurs="0"/>
                                        </xsd:sequence>
                                        <xsd:attribute name="radius" type="xsd:float" use="optional"/>
                                        <xsd:attribute name="transValue" type="transValue" use="optional"/>
                                        <xsd:attributeGroup ref="rgb-color"/>
                                        <xsd:attribute name="id" type="xsd:ID"/>
                                        <xsd:attribute name="start-time" type="xsd:string" use="optional"/>
                                        <xsd:attribute name="end-time" type="xsd:string" use="optional"/>
                                        <xsd:attribute name="duration" type="xsd:string" use="optional"/>
                    </xsd:complexType>
</xsd:element>
```

**Fig. 5.** Example of defined XML schema by XSD.

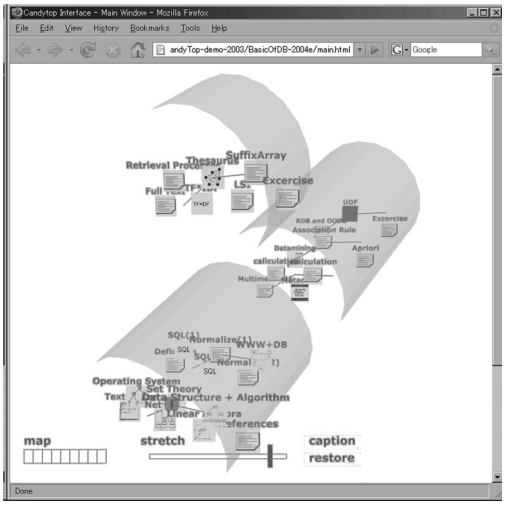


**Fig. 7.** Example of generated CandyTop visualization instance

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<world>
    <candy>
      <drop r="0" b="0" g="0" start-time="0" id="2901">
        <grain start-time="42000" id="2125">
          <position y="0.0" z="0.0" x="42000.0"/>
          <icon-uri>http://media.rsl.crl.fujixerox.co.jp/mediamagic/servlet/ImgServ?id=KEY_FRAME.B_IMG&amp;param=KF_ID=2125</icon-uri>
        </grain>
      </drop>
      <drop r="0" b="0" g="0" start-time="50000" id="2902">
        <grain start-time="50000" id="2126">
          <position y="0.0" z="0.0" x="50000.0"/>
          <icon-uri>http://media.rsl.crl.fujixerox.co.jp/mediamagic/servlet/ImgServ?id=KEY_FRAME.B_IMG&amp;param=KF_ID=2126</icon-uri>
        </grain>
      </drop>
      <drop r="0" b="0" g="0" start-time="51000" id="2903">
        <grain start-time="53000" id="2127">
          <position y="0.0" z="0.0" x="53000.0"/>
          <icon-uri>http://media.rsl.crl.fujixerox.co.jp/mediamagic/servlet/ImgServ?id=KEY_FRAME.B_IMG&amp;param=KF_ID=2127</icon-uri>
        </grain>
      </drop>
      <drop r="0" b="0" g="0" start-time="67000" id="2904">
        <grain start-time="71000" id="2128">
          <position y="0.0" z="0.0" x="71000.0"/>
          <icon-uri>http://media.rsl.crl.fujixerox.co.jp/mediamagic/servlet/ImgServ?id=KEY_FRAME.B_IMG&amp;param=KF_ID=2128</icon-uri>
        </grain>
      </drop>
      <drop r="0" b="0" g="0" start-time="84000" id="2905">
        <grain start-time="86000" id="2129">
          <position y="0.0" z="0.0" x="86000.0"/>
          <icon-uri>http://media.rsl.crl.fujixerox.co.jp/mediamagic/servlet/ImgServ?id=KEY_FRAME.B_IMG&amp;param=KF_ID=2129</icon-uri>
        </grain>
      </drop>
```

**Fig. 6.** Example of generated XML instance.

Schema Definition (XSD):

- Making structure for visualizing elements.
- Defining characteristics for each visualizing element as its attributes

Fig. 5 shows an example of defined XML schema by this step on the method.

### 4.4 Step-2: Compiling of schema

From the XML schema generated by Step-1, we generate a XML processing basic modules of visualization automatically for the basis of Visualization Marshaller and Visualization Geometric Engine shown in Fig. 4.

### 4.5 Step-3: Generation of XML instance by visualization marshaller

In this step, we generate XML instances from media databases for multimedia data by Visualization Marshaller.

A Visualization Marshaller is a sub-module of the method and it is implemented based on the XML processing basic module generated by Step-2. It is designed to generate XML instances automatically according to the XML schema defined by Step-1.

### 4.5 Step-4: Generation of visualization instance by visualization geometric engine

In this step, we generate visualization instances from XML instances by Visualization Geometric Engine.

A Visualization Geometric Engine is a sub-module of the method, and it is implemented based on the XML processing basic module generated by Step-2. It is designed to arrange, decide positions for multimedia data, and generate visualization (Fig. 7) instances automatically to recognize holistic and detail relationships for multimedia data.

## 5. Experiment

In this section, we discuss the feasibility and effectiveness of the visualization method described in the section 4 for the application of e-Learning (Yoshida 2004).[17)]

### 5.1 Experimental environment

We have two lecture for user study to clarify feasibility of effectiveness of holistic and detail relationships (Yoshida 2004).[17)] These lectures are on database systems and its applications.

We have designed and had a first lecture without the visualization, and a second one with the

visualization (Fig. 7). We had 14 testee learners for the first lecture, and 11 learners for the second lecture. University student and people from companies on information technology was included in them. We made conditions of people equal for these two lectures on skills for information technology, experience on database systems.

### 5.2 Experimental methodology

In order to clarify feasibility of effectiveness of holistic and detail relationships recognition of multimeda data, we have two experiments as follows:

- Experiment-A: evaluation of degree of comprehension by learners themselves
- Experiment-B: evaluation of degree of comprehension by examinations

In these experiment, we show that the visualization will increase the degree of comprehension in comparison with the case without the visualization. Experiment-A is the evaluation from learners' point of view, and Experiment-B is the evaluation from objective viewpoint.

We had 41 questions for each learners in Experiment-A. We set four choices for each question. The choice "1" stands for "Well Done" for each question corresponding to each learning element. The choice "2" stands for "Done." The choice "3" stands for "Not Attaining," and the choice "4" stands for "Not Attaining at All."

We had examinations for all learners in both lectures on Experiment-B. We had marked them out of 5 point for each examination. We checked the average point for both case of the lecture without the visualization and with the visualization.

### 5.3 Experimental results

We show the result of Experiment-A as Figs. 8 and 9. Figure 8 shows average point for learners' answers corresponding all 41 questions. Figure 8 shows average point for learners' answers corresponding 10 questions (in 41 questions) only for asking the relationships of learning materials.

We can see the average points of the second lecture (with the visualization) are high in comparison with the first lecture (without the visualization) for all questions by Fig. 8. The average point for answer "1" and "2" ("Well Done" and "Done") is high in comparison with the points for negative answers.

We can see the average points of the second lecture (with the visualization) are higher than the average points of first lecture (without the

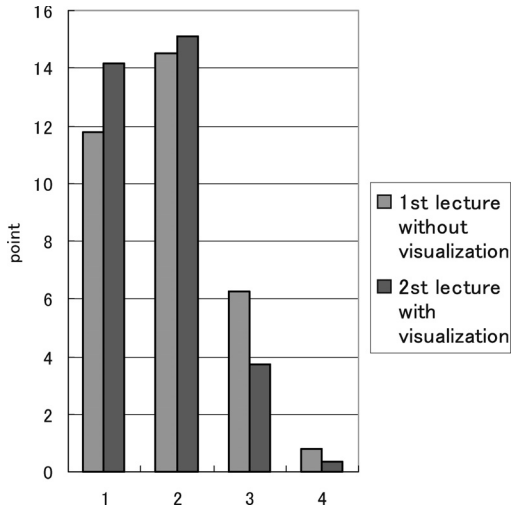


**Fig. 8.** Points (degree of comprehension) of learners for evaluation (average values for whole questions) 1: average point for answers "Well Done," 2: average point for answers "Done," 3: average point for answers "Not Attaining," 4: average point for answers "Not Attaining at All."

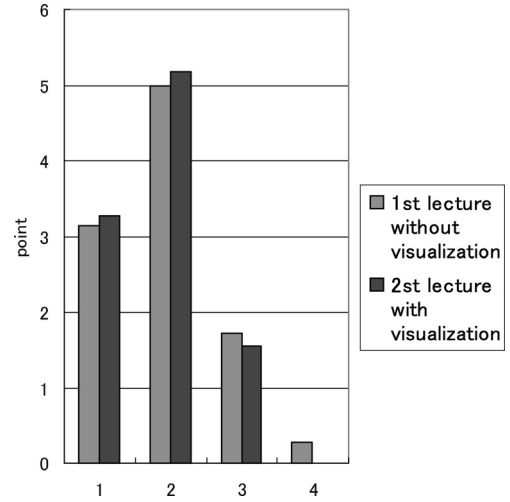


**Fig. 9.** Points (degree of comprehension) of learners for evaluation (average values for questions corresponding relationships of educational materials) 1: average point for answers "Well Done," 2: average point for answers "Done," 3: average point for answers "Not Attaining," 4: average point for answers "Not Attaining at All."

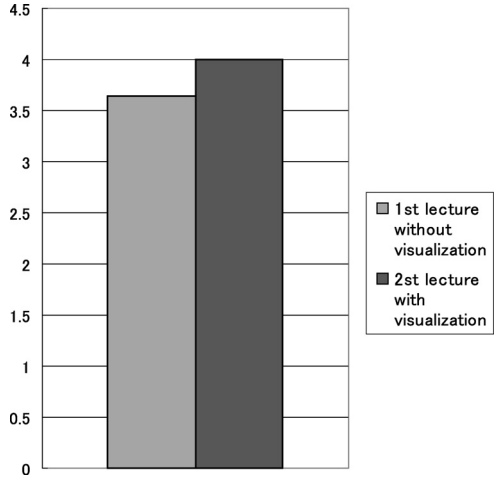


**Fig. 10.** Average scores of examinations in both lectures.

visualization) for 10 questions only for asking the relationships of learning materials by Fig. 9.

The average point for answer "1" and "2" ("Well Done" and "Done") is high in comparison with the points for negative answers in both case of counting 41 all questions and only 10 questions on asking the relationships of learning materials.

We show the result of Experiment-B as Fig. 10.

We can see the average score of examination on the second lecture is higher than that of the first lecture. We can say that the score is higher with the visualization than the score without the visualization.

We can say that the visualization is effective from the viewpoint of learners themselves by Experiment-A as Figs. 8 and 9.

Especially, Fig. 9 mentioned the important point that the average point with the visualization is much higher than that without the visualization on asking only 10 questions on the relationships of learning materials. This shows the direct effectiveness of the visualization method on improving comprehension of relationships for learning materials.

We can say that the visualization is effective from the objective viewpoint by Experiment-B as shown Fig. 10.

These experimental results shows the feasibility of the implementation method of automatic generation for visualization on e-Learning environment. And the results also shows the effectiveness of the visualization for learning materials.

## 6. Conclusion

In this paper, we discuss on multimedia database technologies especially for a metadata extraction method for video streams, a retrieval method by keywords and a given strategies, and a data mining method by 3D visualization method and its application to e-learning.

The main features of the methods are: (1) immediate and automatic extraction of metadata by giving semantics to combinations of heterogeneous sensors for video streams, (2) flexible retrieval according to the selected strategy from broadening, deepening and expanding, for metadata of video streams, (3) automatic 3D visualization for holistic and detail relationships of video streams.

We also discuss its feasibility by experimental results.

As our future work, retrieval methods by impressions or emotional words, data mining methods for document data, quantitative analysis by large databases for each method must be realized.

## Acknowledgement

This research activity is a collaborative work with Fuji Xerox Co., Ltd.

### References

1) M. Brady: "Computational Approaches to Image Understanding," ACM Computing Surveys (CSUR), Volume 14, Issue 1, pp. 3–71, 1982.
2) J. Lamping and R. Rao: "The Hyperbolic Browser: A Focus+Context Technique for Visualizing Large Hierarchies," Journal of Visual Languages and Computing, Vol. 7, No. 1, pp. 33–55, 1996.
3) J. D. Mackinlay, G. G. Robertson and S. K. Card: "The Perspective Wall: Detail and context smoothly integrated," Proc. of International Conference of CHI'91, ACM Press, pp. 173–179, 1991.
4) M. F. McTear: "Spoken dialogue technology: enabling the conversational user interface," ACM Computing Surveys (CSUR), Vol. 34, Issue 1, pp. 90–169, 2002.
5) J. Han and M. Kamber: "Data Mining: Concepts and Techniques," Morgan Kaufmann Pub., 550 pp., 2000.
6) A. Jaimes, N. Yoshida, K. Murai, K. Hirata and J. Miyazaki: "Interactive Visualization of Multi-Stream Meeting Videos based on Automatic Visual Content Analysis," In Proceedings of 2004 IEEE International Workshop on Multimedia

Signal Processing, Sep. 2004.
7) S. Khoshafian and B. Baker: "Multimedia and Imaging Databases," Morgan Kaufmann, ISBN: 1558603123, 586 pp., 1995.
8) G. G. Robertson and J. D. Mackinlay: "The Document Lens," Proc. ACM UIST'93, pp. 101–108, 1993.
9) A. Sheth and W. Klas (eds.): Multimedia Data Management—using metadata to integrate and apply digital media, McGraw Hill, 1998.
10) N. Shimizu, J. Nakamura, N. Yoshida, T. Hattori and T. Hagino: "Personalization of Materials for Learning on Demand Using RDF," Proceedings of International Conference of WWW 2003, May 2003.
11) T. Strothotte and H. Wagener: "Computational Visualization: Graphics, Abstraction, and Interactivity," Springer Verlag, 1999.
12) S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky: "Video Manga: Generating Semantically Meaningful Video Summaries," Proceedings ACM Multimedia, ACM Press, pp. 383–392, 1999.
13) U. Westermann and W. Klas: An analysis of XML database solutions for the management of MPEG-7 media descriptions, ACM Computing Surveys (CSUR), Volume 35, Issue 4, December 2003.
14) A. Yilmaz, O. Javed and M. Shah: "Object tracking: A survey," ACM Computing Surveys (CSUR), Volume 38 , Issue 4, pp. 1–45, 2006.
15) N. Yoshida, J. Miyazaki and A. Wakita: "CandyTop Interface: A Visualization Method with Positive Attention for Growing Multimedia Documents," Proc. of 7th International Conference on Information Visualization (IV03), published by IEEE Computer Society, London, UK, July 2003.
16) A. Wakita. N. Yoshida, J. Miyazaki and H. Chiyokura: "Candytop : A Web3D Interface to Visualize Growth of Multimedia Documents," Proceedings of the 30th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH2003), Sandiego, California, USA, July 2003.
17) N. Yoshida, K. Hirata and J. Miyazaki: "An Automatic Generation Method of 3D Visualization for Holistic and Detail Relationships on e-Learning Environment," Proceedings of International Workshop on Cyberspace Technologies and Societies (IWCTS2004), in conjection with the 2004 International Symposium on Applications and the Internet (SAINT2004), Jan. 2004.
18) N. Yoshida and J. Miyazaki: "An Automatic and Immediate Metadata Extraction Method by Heterogeneous Sensors for Meeting Video Streams," IEEE International Symposium on Applications and the Internet (SAINT 2005)—the International Workshop on Cyberspace Technologies and Societies (IWCTS 2005), pp. 446–449, IEEE Computer Society Press, Feb. 2005.