

AtomsNet: Multimedia Peer2Peer File Sharing

Willem de Bruijn and Michael S. Lew

LIACS Media Lab, Leiden University
Niels Bohrweg 1, 2300 CA Leiden, The Netherlands
{wdebruijn, mlew}@liacs.nl

Abstract. Current Peer2Peer (P2P) systems such as Napster or Kazaa do not perform analysis on the content of the media but instead depend on manual text annotation. In the AtomsNet project we are investigating multi-modal content based browsing and searching methods for P2P retrieval systems. This is the first P2P system which performs analysis on the video content for browsing multimedia collections over large, distributed P2P networks.

1 Introduction

Peer2Peer (P2P) networks have had major attention recently and typically have millions of peers within their sub-network. The widely popular applications designed specifically for sharing information amongst peers have only emerged a few years ago. The best known example of P2P networks is the music sharing program called Napster (www.napster.com).

The strongpoint of Napster is very clear. Many more songs are available for free using this network than anywhere else. Live recordings, bootlegs and "rare takes" were all made available by music enthusiasts. Unfortunately Napster suffered from several weaknesses: legal problems and the difficulty in searching for files - the user had to search through an unordered bag of files.

Napster's shortcomings can be dealt with in several ways. We started this project in 2000, when no alternative to the Napster network had fully emerged. Our intention was to explore the possible solutions to the more general problem of multimedia retrieval. Since then many different P2P applications have been created, such as Kazaa, Morpheus and Gnutella. Taking into account the current state of the art, we will try to address which methods can be used to improve the aspects of multimedia retrieval over large scale P2P networks.

2 Background

The number of people using the internet has grown rapidly over the last couple of years. With this increase in public involvement a shift occurred in networking methodology. Traditionally, there existed a large gap between the suppliers and the consumers of information. In the popular client/server paradigm suppliers would be

active all of the time, servicing requests from clients at a fixed location. In an environment where many participants are both consumers and producers of information and no server can be trusted to be active, however, another way of communicating is necessary.

With the recent growth of available information on the internet and the lack of internal structure linking this data, a way of indexing information based not on hierarchy but on equality has to be found. The current situation renewed interest, both from the scientific community and from the general public, in distributed networks. Distributed networks, also referred to as peer-to-peer or P2P networks, index information in a decentralized manner, thus acknowledging the fact that no internal hierarchy exists.

There are other interesting WWW image search engines which have been described in the research literature. The WebSeek[17] system from Columbia University finds similar images and performs automatic text based category classification. The Webseer[3] system from the University of Chicago lets users search by the number of faces and by text queries. Taycher, Cascia, and Sclaroff[19] designed the ImageRover system to primarily use relevance feedback for the search process, and in the PicToSeek system, Gevers and Smeulders[4] search through the images using similar images and image features. Lew [10] describes one of the first web systems which finds images and video based on understanding of semantic categories.

In visual information retrieval there has been relevant previous research [1-22]. Picard[5] reported promising results in classifying blocks in an image into "at a glance" categories. What this means is that she investigated multiple model methods to classify an NxN block into categories which humans could classify without logically analyzing the content. Forsyth, et al. [2] found objects from feature blobs. More recently, Vailaya, Jain, and Zhang [21] have reported success in classifying images as city vs. landscape. They found that the edge direction features have sufficient discriminatory power for accurate classification of their test set. The commonality between these methods was using multiple features for object/concept detection. Regarding object detection, the recent surge of research toward face recognition has motivated robust methods for face detection in complex scenery. Representative results have been reported by Sung and Poggio[18], Rowley and Kanade[16], Lew and Huijsmans [11], and Lew and Huang[12].

When searching for an item one always has to describe certain characteristics that distinguish the desired item from others. These characteristics will then be cross-checked with the characteristics of the possible options to select the best answer. It doesn't matter if a person is searching for a house by asking directions or if he is trying to find a digital document on the internet, characteristic negotiation is the basic element of searching.

When dealing specifically with data, these characteristics can be seen as data describing the data. For this special type of information the term `{it metadata}` is generally used, where meta stands for 'above'. In everyday situations this information is completely distinct from the object it details. When dealing with digital information, however, it all boils down to numbers. This means that metadata can be seen both as a description and an object itself.

Traditionally metadata has been separated from objects. The most widely used form of metadata is the filesystem structure used to index digital data on a harddrive. This data is identical for every file and therefore extremely brief. Such an

implementation might be sufficient for a local filesystem where a small group of people handle the data, but it is an impractical indexing method for use in large heterogeneous environments. The large scale on which information is made available on the internet increase the demand for automation of resource identification to a level beyond that of basic filesystems.

Currently a few projects are underway to create a more flexible framework for describing data. Since standardization is essential for any system to become widespread the initiatives backed by international organizations have the highest chance of success. The Resource Description Format backed by the influential World Wide Web consortium, is a semantic layer placed on top of the meaningless XML syntax. In RDF, meaning is given to data by coupling RDF formatted content to specialized dictionaries. It is already being used to identify and rebroadcast news messages on the internet and by the Open Directory Project webdirectory. RDF has been designed specifically with the automation of internet resource discovery in mind. The developers expect RDF to become the framework underlining the so called Semantic Web, a global information network where computers can autonomously index information.

Despite the effort of various interest groups, metadata is for the larger part currently written in proprietary formats, eliminating interoperability between software systems. Future developers should try to use standardized frameworks as the RDF mentioned above as much as possible if the internet is to remain a global information domain.

Here we summarize the de facto standards and innovative implementations:

- **Napster** and **Gnutella** - mostly music sharing; parsing mp3 headers for text annotation.
- **KaZaa** and **Morpheus** - filesystem info + manually entered info that is saved between transfers
- **AudioGalaxy** - searching using knowledge about user behaviour (link multiple downloads from the same user as potentially connected, thus creating subcategories based on user).

Note that none of these systems uses content based analysis in the browse/search process. In fact, manual text annotation is a critical factor in all of their search engines. This has also led to many users download viruses because some malicious users used false text annotation for their viruses.

3 AtomsNet: Network Design and Content Based Analysis

AtomsNet was designed to be scalable to the entire Internet, which currently means approximately 700,000 networked computers. We describe the design of the interconnection network and the content based analysis in this section. The AtomsNet system allows the following query types:

- text keywords
- category based searching

It provides functionality to

- automatically categorize all of the files in a directory
 - you can drag and drop local directories for automatic analysis
- allow plug-ins for custom media analysis and browsing:
 - extract text annotation from mp3 files
 - smart text searches: "find videos starring Harrison Ford"
 - extract representative keyframes from MPEG video

3.1 How Do You Find Your Peers?

Since the Web is not an organized database, it is challenging if not impossible to comprehensively find all of the hyperlinks. The general procedure used by typical search engines is to begin with an initial set of hyperlinks, follow each hyperlink to a page, parse each page for new hyperlinks, and follow the new hyperlinks, repeat ad infinitum. This typically results in a breadth first search of the WWW as shown in Figure 1. It also means that the comprehensiveness of any search engine is limited to the sites which are reachable from the initial set of seed links. For example, if there are no hyperlinks connecting networks in the U.S. and China, then a search engine which begins from exclusively U.S. links will never find the network in China.

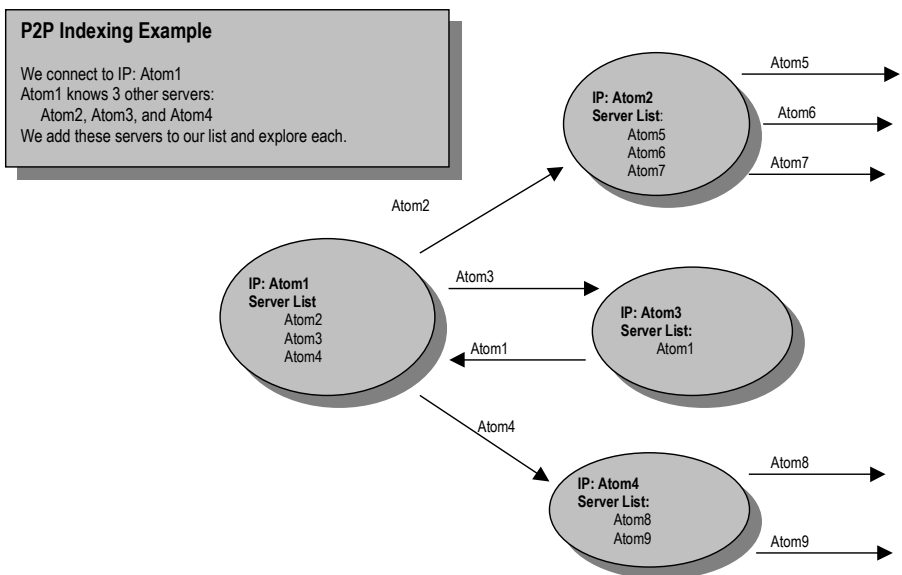


Fig. 1. P2P indexing example

In our scheme, AtomsNet is composed of atoms which are the peer computers. The typical atom is expected to have a slow connection of roughly 56Kbit/s to 1 Mbit/s. Atoms are connected in the virtual network by SuperAtoms which typically have 1 Mbit/s or greater network bandwidth.

In designing AtomsNet distributed network, the critical factor was to minimize the communications overhead between the computers. Our approach was to incorporate multiple tier hierarchical routing into the P2P search system. At the lowest tier, each atom was given sole responsibility for a list of hosts. Each atom sends its list of atoms to the next higher tier.

When an atom connects to the AtomsNet, it contacts a superatom for the list of active atoms and superatoms. The list of atoms and superatoms are sorted by topological network distance. When the atom posts a query, it is only searching the local list of atoms and superatoms. The superatom relays this information periodically to all of the child atoms. Hierarchical routing has the significant advantage that it is extendable to arbitrarily large networks as long as the bandwidths are sufficient.

3.2 Content Based Analysis

When processing a video, the client machine should have a minimal workload in P2P situations. Most video abstraction and summarization methods entail significant computational complexity which would result in the client machine being unusable for minutes or even hours.

Our goal is to extract a single frame from the entire video to represent the content. The reasoning for this is that each result will also return a pictorial keyframe so low bandwidth users presumably would not want dozens or hundreds of keyframes. A secondary goal is to be able to extract the representative keyframe at a rate of 3 full length MPEG movies (approx. 1GB per movie) per second.

There is only one work which is similar to our direction. Dufaux[15] implemented a method for extracting a single keyframe from a video by integrating motion and spatial activity analysis, which results in a single keyframe for the entire video. His process was computationally expensive (minutes per video) and not a possibility for our P2P application.

In order to achieve the desired keyframe extraction rate, some compromises clearly need to be made. First, we can not even look at every frame in the movies because it would require too much time - a gigabyte movie would take 100 seconds to read on a hard disk capable of a sustained 10 MB/s access.

Therefore, we turned to sampling N MPEG I-frames which are in the second/third quarter of the video. We do not process the beginning of the video because many movies have less relevant "setup" scenes in the early shots.

Here we make a pure heuristic. Subjectively, we assert that keyframes which contain people are more interesting than ones that do not. There have been several methods for detecting faces (see Rowley, et al. [16]) in images with complex backgrounds, but these are also typically computationally complex.

Thus, we design a simple face detector using color. Methods in the literature have used the well known color spaces for skin detection. Our novel contribution is to design a new color space specifically for skin detection.

In principle, we do not care about representing skin, but only classifying a pixel as skin color based on the pixel and a small region around it. The optimal linear classifier can be expressed as

$$\text{maximize } \frac{k^t \sum_x k}{k^t \sum_n k} \quad (1)$$

where k is the vector containing the pixel and its neighbors. We do not go into detail about equation (1) because it is found in most pattern recognition books. Instead we explain how we use it. The process is that we create training sets which comprise small skin regions and small non-skin regions. We find the orthogonal basis which gives us the optimal linear classifier which satisfies equation (1). This basis is the new color space for optimally classifying color pixel regions as skin or non-skin.

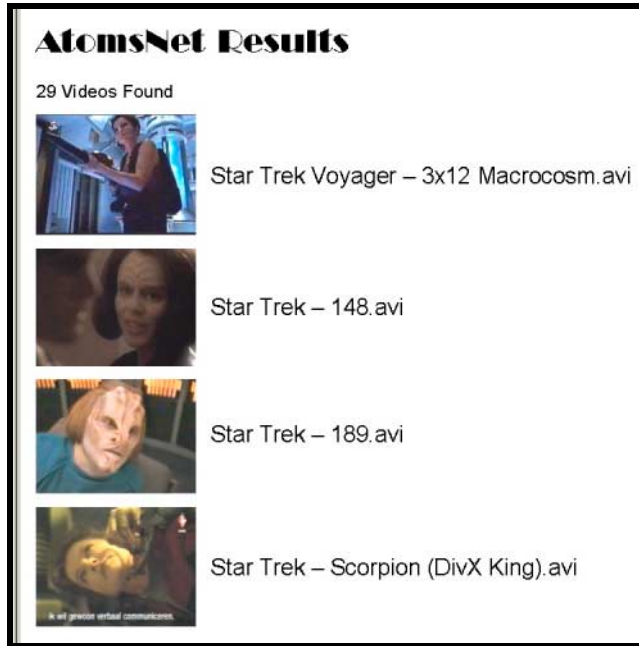


Fig. 3. An example of results from AtomsNet

A minimum distance classifier is used for deciding between skin/non-skin. After we classify each pixel as skin or non-skin, we select the keyframe which has the largest percentage of skin in it as shown in Figure 2.

For a PIII/800Mhz with a Seagate Cheetah Ultra160 73 GB harddrive, our system can analyze 9 keyframes per video and reach the goal of processing 3 full length videos per second. An example of the results page from AtomsNet for a video search on "star trek" is displayed in Figure 3.

4 Conclusions

In the AtomsNet multimedia P2P system, we needed to satisfy network considerations and allow browsing based on the content of the video. The main contribution of this paper toward video retrieval is the heuristic algorithm for finding a single keyframe to represent an entire video. It was designed intentionally to minimize computational complexity. The novel aspect of the keyframe selection algorithm was to create a new color space based on the theory of optimal linear classifiers. This new color space is optimized for classifying small pixel regions into the categories of skin or non-skin.

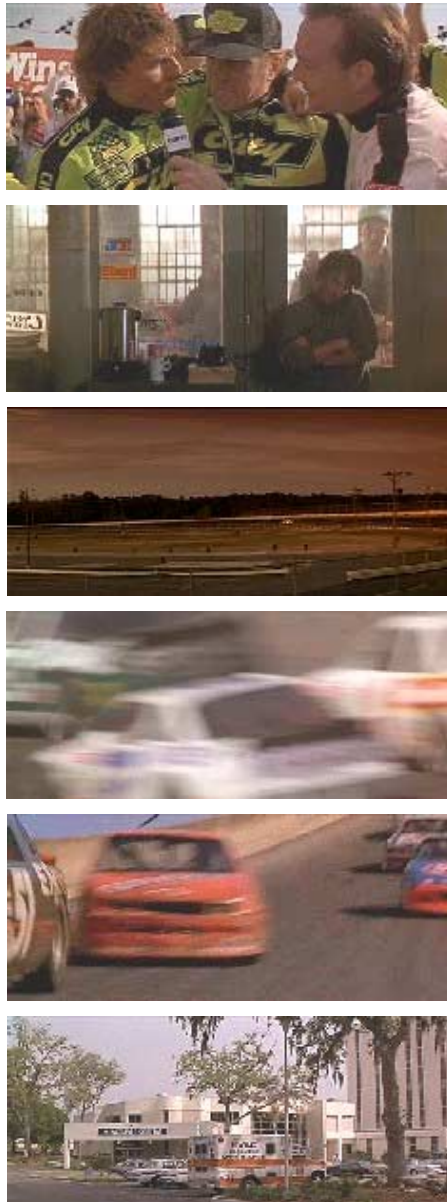


Fig. 2. Sampled frames ordered by most skin to least skin

There is ample room for improvements in P2P networks. The networking aspect of finding peers remains as a difficult problem. Browsing and searching for different media types are also major challenges. One advantage of the AtomsNet framework is that plugins can be created for each type of media. In the near future, our intention is to focus on the video analysis aspect toward extracting semantic descriptive information.

Acknowledgments

This project was assisted and supported by Altavista (Cambridge Research Lab), Magicbot, and the LIACS Media Lab. The authors would like to thank Michael Swain for beneficial and interesting discussions on searching the internet.

References

- [1] Del Bimbo, A., and P. Pala, "Visual Image Retrieval by Elastic Matching of User Sketches," *IEEE Trans. Pattern Analysis and Machine Intelligence*, February, pp. 121-132, 1997.
- [2] Forsyth, D., J. Malik, M. Fleck, T. Leung, C. Bregler, C. Carson, and H. Greenspan, "Finding Pictures of Objects in Large Collections of Images," *Proceedings, International Workshop on Object Recognition*, Cambridge, April 1996.
- [3] Frankel, C., M. Swain and V. Athitsos, "WebSeer: An Image Search Engine for the World Wide Web," *Technical Report 96-14*, University of Chicago, August 1996.
- [4] Gevers, T. and A. Smeulders, "PicToSeek: A Content-Based Image Search System for the World Wide Web," *VISUAL'97*, San Diego, December, pp. 93-100.
- [5] Picard, R. "A Society of Models for Video and Image Libraries." *IBM Systems Journal*. 1996.
- [6] Hu, M., "Visual Pattern Recognition by Moment Invariants", *IRA Trans. on Information Theory*, vol. 17-8, no. 2, pp. 179-187, Feb. 1962.
- [7] Huijsmans, D. P., M. Lew, and D. Denteneer, "Quality Measures for Interactive Image Retrieval with a Performance Evaluation of Two 3x3 Texel-based Methods," *International Conference on Image Analysis and Processing*, Florence, Italy, September, 1997.
- [8] Kittler, J., M. Hatef, R. Duin, and J. Matas, "On Combining Classifiers," *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 20, no. 3, March 1998.
- [9] Kullback, S. "Information Theory and Statistics," Wiley, New York, 1959.
- [10] Lew, M., "Next Generation Web Searches for Visual Content," *IEEE Computer*, November, pp. 46-53, 2000.
- [11] Lew, M. and N. Huijsmans, "Information Theory and Face Detection," *Proceedings of the International Conference on Pattern Recognition*, Vienna, Austria, August 25-30, 1996, pp.601-605.
- [12] Lew, M. and T. Huang, "Optimal Supports for Image Matching," *Proc. of the IEEE Digital Signal Processing Workshop*, Loen, Norway, Sept. 1-4, 1996, pp. 251-254.
- [13] Ojala, T., M. Pietikainen and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions," vol. 29, no. 1, pp. 51-59, 1996.
- [14] Petkovic, D., "Challenges and Opportunities for Pattern Recognition and Computer Vision Research in Year 2000 and Beyond," *Proc. of the Int. Conf. on Image Analysis and Processing*, September, Florence, vol. 2, pp. 1-5, 1997.

- [15] Dufaux, F. "Key Frame Selection to Represent a Video." ICIP, 2000.
- [16] Rowley, H., and T. Kanade, Neural Network Based Face Detection, IEEE Trans. Patt. Anal. and Mach. Intell., vol. 20, no. 1, pp. 23-38, 1998.
- [17] Smith, J. R. and S. F. Chang, "Visually Searching the Web for Content," IEEE Multimedia, 1997, pg. 12-20.
- [18] Sung, K. K., and T. Poggio, Example-Based Learning for View-Based Human Face Detection, IEEE Trans. on Patt. Anal. and Mach. Intell, vol. 20, no. 1, pp. 39-51, 1998.
- [19] Taycher, L., M. Cascia, and S. Sclaroff, "Image Digestion and Relevance Feedback in the ImageRover WWW Search Engine," VISUAL'97, December, San Diego, pp. 85-91.
- [20] Tekalp, A. M., Digital Video Processing, Prentice Hall, New Jersey, 1995.
- [21] Vailaya, A., A. Jain and H. Zhang, "On Image Classification: City vs. Landscape," IEEE Workshop on Content-Based Access of Image and Video Libraries, Santa Barbara, June 21, 1998.
- [22] Wang, L. and D. C. He, "Texture Classification Using Texture Spectrum," Pattern Recognition 23, pp. 905-910, 1990.