CLIMS - A system for image retrieval by using colour and wavelet features

O. Kao, I. la Tendresse

Department of Computer Science, Technical University of Clausthal
Julius-Albert-Strasse 4, D-38678 Clausthal-Zellerfeld, Germany

Abstract. In this paper a system called CLIMS (*CLausthal Image Management System*) for content based image retrieval as an important subsystem of a general multimedia database is presented. It offers querying by sketch and image example and uses colour and wavelet based features for the comparison of images. Each image in the database is represented by a set of wavelet coefficients and colour attributes, which form the fundament for the retrieval.
In order to enable efficient similarity search two index structures, VP-Trees and L^q metric, are introduced and discussed. With the extension of the original VP-tree algorithm a ranking of the n most similar images is possible. The efficiency of the proposed retrieval methods is evaluated on a sample, general image catalogue.

1 Introduction

The development of the information technology in the 1990s is often described as a multimedia revolution. Multimedia is the synchronised association of time dependent (dynamic) and time independent (static) media. Dynamic media are audio and video streams as well as animation. Text, graphics and images are part of the static media. Although the hardware (I/O devices, memory, networks etc.) and software infrastructure have improved significantly 90% of the information is still on paper. One of the reasons is the lack of reliable methods for content analysis of the different media types, for example we still do not have suitable image segmentation and analysis algorithms. Furthermore basic mechanisms and technologies for the realisation of multimedia database management systems, e.g. information retrieval are not available. On the other hand subsystems like CD-ROMs with multimedia content, thumbnail systems, video on demand etc. can be seen as a part of a multimedia database [1]. Many existing database systems handle multimedia objects as BLOBs (*Binary Large Objects*). Each object is characterised by a manually compiled set of key words, thus content retrieval bases on full text search in this set. Problems occur due to the disadvantageous reduction of the complex media content on a small number of key words.

Therefore methods for content based information retrieval are necessary. In this paper a system called CLIMS (*CLausthal Image Management System*) for content based image retrieval as an important subsystem of a multimedia database is discussed.

2 Image databases

An image database is a system for the archival and retrieval of images and can be found in many areas, for example medical applications, remote sensing, news agencies, authorities, museums etc. It differs from systems for pattern recognition: in such systems the number of possible patterns is small, the accuracy high and the recording conditions are not significantly changing. The results are usually transferred to a control device and automatically evaluated.

In opposition thereof an image database contains a large set of images from different classes, which are taken under various conditions. The goal is to find a number of images, which are similar to a given sample image. High recognition rates are desired but not always possible. Furthermore the system response time is much longer than in the case of the pattern recognition. For the extraction and organisation of the image information suitable analysis and knowledge discovery methods are necessary. A work around of these unsolved problems is realised with the idea of content based image retrieval (CBIR). It is a class of search algorithms based on the extraction and comparison of low level image features, which are combined in logical features in order to represent the image content on a higher abstraction level.

The features can be extracted manually or automatically. In the first case the user marks the interesting areas with sketch tools. Subsequently all images in the database are processed in order to find similar regions. A reasonable system response time is possible if only a small number of images is considered. This is the reason, why many existing systems use a-priori extracted features. These are calculated, when the image is inserted into the database and compared with the corresponding features of the query image.

An important difference between image and conventional databases is given by the querying mechanisms. Languages such as SQL are not very efficient, because the image features are usually abstract numbers and arrays and thus not understandable to users without image processing background. Therefore visual querying methods like query-by-pictorial-example (QBPE), query-by-painting or sketch retrieval [2, 3] are preferred. The user creates an example sketch or loads an example image, which are subsequently processed in the same manner as the images in the database.

3 CLIMS

The image database CLIMS is an experimental system for content based retrieval in general image catalogues. The system consists of clear defined modules, which can be easily extended with new methods for feature extraction and retrieval. The CLIMS basic components are the image processing system CLIPS, the relational database system PostgreSQL [4] and a web-based user interface.

CLIPS contains a selection of standard image processing methods such as colour transformations, segmentation etc. Each of this operators is realised as an independent UNIX shell program: it recieves the image data via the standard input and writes the results on the standard output. For a suitable integration of the available filter in the image database the user interface was re-designed and consists now of two web-based modules, which are shown in Figure 1. The draw module offers a selection of basic functions for creation and modification of graphic primitives like lines, ellipses, rectangles etc. In addition to the query by sketch the user can load images in GIF or JPEG format as a starting point for the similarity based search. The browser is used for the visualisation of the query results.

Fig. 1. Graphical user interface: sketching tools and browser for the retrieval results

Once the sample image/sketch is created, the query is started on one of the implemented colour models. Each model offers different advantages and disadvantages for the retrieval of certain image classes, thus further investigations are necessary in order to determine most suitable combinations. We used the PostgreSQL C programming library for the implementation of the functions for the communication with the database server. The basic operations are creation of a new image catalogue, insertion of features, index structures and images and content based search for the n most similar images by using one of the available features.

For each image entered in the database unique identifier and a table column for the values of all available features are created. Subsequently the algorithm for the feature extraction is implemented as a CLIPS filter, tested and finally applied on all images in the database. The calculated results are inserted into the corresponding tables and, if possible, into the index structures, for example

VP-trees (see section 5). These support the similarity search and reduce the computational complexity for the creation of the image ranking.

In the current version of CLIMS two algorithms for feature extraction—based on colour values and wavelet coefficients—are available. The application of the algorithms on the sample image or sketch produces the search parameter. With these values the corresponding index structures, which contain the already calculated values for all images in the database, are searched resulting into a ranking of the n most similar images, where n is a constant value given by the user. In the last step the image raw data is determined, loaded and sent to the user interface for visualisation. Following section presents a detailed discussion of the extraction algorithms.

4 CLIMS image features

Colour based features are often used for the comparison of images. A histogram defines e.g. the colour distribution over the image, thus images having similar histograms are considered similar. This method delivers good results, if all examined images belong to the same class, for example landscape images. Otherwise this information is not sufficient for general retrieval.

Another disadvantage is given by the large time and computing effort for the calculation and comparison of histograms in the case of true colour images, which have up to 16,7 Million colours. Therefore such images are pre-processed in order to reduce the number of colours. The most simple algorithm is the fix conversion: the amount of bit per colour channel is decreased. More suitable conversion can be achieved with adaptive methods like the Octree-Downfilter or Median Cut algorithm. The resulting histograms can not be compared directly, because each component of the histogram may represent a different colour. Therefore an assignment of the determined to a given set of reference colours is necessary.

Another colour based approach for image comparison uses statistical colour attributes introduced by Stricker and Orengo in [5]. The complex functions for the approximation of the image histogram are replaced by single values like the average E_i, the variance σ_i and the asymmetry s_i. The definition of these elements for the i-th colour channel and the j-th pixel of an image containing N pixel is given by the following Equations:

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij} \tag{1}$$

$$\sigma_i = (\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^2)^{\frac{1}{2}} \tag{2}$$

$$s_i = (\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^3)^{\frac{1}{3}}. \tag{3}$$

For an image in the RGB colour model the size of these features amounts to three floating point values per channel.

Wavelet coefficients can also be used as features for measuring similarity of two pictures. Every one or two dimensional signal can be represented by wavelet coefficients whose number depends on the subdivision process. This process starts with the highest resolution in the first stage and concludes with the lowest resolution in the last stage. Large wavelet coefficients and/or large differences in the high resolution stage can be compared with high frequencies of the original signal. Small differences in a distinct stage can analogously be a sign for a small contribution of a corresponding frequency in the original signal. Two signals can be matched in the final analysis by consideration of the coefficients on every resolution stage. Specific frequency ranges may be ignored or included in the matching procedure with different weights.

A *Haar-Wavelet-Transformation* is defined as following. Let s_0 be a signal with $2^n, n \in \mathbb{N}$ discrete values $s_{0,k}$:

$$s_0 = \{s_{0,k} \mid 0 \leq k < 2^n\}. \tag{4}$$

If the "mean value/difference" transformation is applied on every pair $a = s_{2k}$ and $b = s_{2k+1}$ a total of 2^{n-1} $(k = 0 \ldots 2^{n-1})$ pairs have to be stored:

$$\begin{aligned} s_{1,k} &= (s_{0,2k} + s_{0,2k+1})/2 \\ d_{1,k} &= s_{0,2k+1} - s_{0,2k} \end{aligned} \tag{5}$$

The original signal s_0 with 2^n values is cut up into two signals: s_1 of the mean values $s_{1,k}$ with 2^{n-1} values and d_1 of the differences $d_{1,k}$ with 2^{n-1} values.

The signal s_1 represents the signal s_0 in a coarser resolution. The difference signal d_1 contains the information for the restoration of the original signal s_0 from s_1. The same transformation can be applied to the signal s_1 resulting into s_2 as a new mean signal and d_2 as a new difference signal. This procedure can be performed n times. At the end of the transformation n detailed signals d_j, $0 \leq j \leq n-1$ with 2^{n-j} coefficients each and one signal s_n on the coarsest resolution stage are available. s_n consists of only one value $s_{n,0}$ which corresponds to the mean value of the entire original signal s_0. This value is called the *DC-component* or the *zero frequency* of the signal.

The wavelet subdivision uses the *haar* functions $h_{j,k}$, $j, k \in \mathbb{Z}$, which are defined as

$$h_{j,k}(x) = \begin{cases} 2^{-j/2} & \left| \; 2^j(k-1) < x. < 2^j(k-1/2) \right. \\ -2^{-j/2} & \left| \; 2^j(k-1/2) \leq x < 2^j k \right. \\ 0 & \left| \; \text{otherwise.} \right. \end{cases} \tag{6}$$

The areas of interest of the *haar* functions $h_{j,k}(x)$ are in range $I_{j,k}$:

$$I_{j,k} = [2^j(k-1), 2^j k]. \tag{7}$$

The characteristic function $\chi_{I_{j,k}}(x)$ with $|I_{j,k}|^{-1/2}$ is used as a weighting function

$$\chi_{I_{j,k}}(x) = \begin{cases} |I_{j,k}|^{-1/2} & \left| \; x \in I_{j,k} \right. \\ 0 & \left| \; \text{otherwise.} \right. \end{cases} \tag{8}$$

where $|I_{j,k}|$ is the length of the range $I_{j,k}$. For a given function $f \in L^2(\mathbb{R})$ and a range $I \in \mathbb{R}$, the *haar* coefficient d_I

$$d_I = \int_{-\infty}^{+\infty} f(x) h_I(x) \mathrm{d}x, \tag{9}$$

and the average s_I

$$s_I = \int_{-\infty}^{+\infty} f(x) \chi_I(x) \mathrm{d}x, \tag{10}$$

of f are defined. The numeric computation of the *haar* coefficients necessities $N = 2^n$ discrete values from the function f:

$$\begin{aligned} s_{0,k} &= \sqrt{N} \int_{2^{-n}(k-1)}^{2^{-n}k} f(x) \mathrm{d}x \qquad (11) \\ d_{1,k} &= \frac{1}{\sqrt{2}} (s_{0,2k-1} - s_{0,2k}) \\ s_{1,k} &= \frac{1}{\sqrt{2}} (s_{0,2k-1} + s_{0,2k}) \end{aligned}$$

We apply this procedure n times according to a pyramid diagram. The computation of the coefficients

$$d_{j+1,k} = \frac{1}{\sqrt{2}} (s_{j,2k-1} - s_{j,2k}) \tag{12}$$

and mean values

$$s_{j+1,k} = \frac{1}{\sqrt{2}} (s_{j,2k-1} + s_{j,2k}) \tag{13}$$

for $j = 0, \ldots, n-1$ and $k = 1, \ldots, 2^{n-j-1}$ requires $2(N-1)$ additions and $2N$ multiplications. Full descriptions and derivations of the specified Equations can be found in [3, 9–11].

Before the wavelet transformation is applied each image is scaled to a size of 128×128. Fast wavelet algorithms exist for $2^m \times 2^m$, $m \in \mathbb{N}$ images thus a significant speedup can be achieved. The *haar* basis has a further advantage because it is suitable for the representation of large single coloured segments, which are often contained in the sample sketches. The L^q metric, described in the next section, is used for the determination of the nearest neighbours.

5 Index structures

CLIMS supports VP trees (*Vantage point*) as an index structure for content based retrieval and L^q metrics for the comparison of vectors with wavelet coefficients. VP trees are introduced by CHIUEH in [6] and enable an efficient next neighbour search. For the application in our image database we extended the algorithms, basically a few special cases have to be analysed (see [7]), thus a ranking of the $k > 1$ next neighbours can be generated.

Let M denote a set of n-dimensional vectors with image features as components. In the inside nodes of a VP tree the vantage points are stored. The leaf nodes contain the values of the extracted features. The distribution of these values to the vantage point can be easily demonstrated, if a binary relation is given: all values are sorted according to the distance to the—randomly chosen—vantage point. Let μ be the median of all distances, thus each sub tree, $S_>$ and $S_\leq$ contains half of all points. Both sub trees have to be searched if the distance of the search parameter p to the vantage point v is in the interval $[\mu - \sigma, \mu + \sigma]$.

The next neighbour search starts in the root node by calculating the distance of the search parameter q to the vantage point v and σ is the—estimated—maximum distance. The next neighbour is element of a subset, if following holds:

$$\mu_i - \sigma < d(q, v) \leq \mu_{i+1} + \sigma. \tag{14}$$

All subsets fulfilling this criterion are examined until the next neighbour is found. An unsuccessful search occurs if the value σ is too small, thus the whole process has to be repeated with a modified σ. This adaptation can be performed by addition or multiplication of constant value.

For the evaluation of the image similarity based on the wavelet feature only the m largest coefficients are considered. All images in the database are expanded or reduced to a standard size of 128×128, thus 16384 coefficients are possible. In the current implementation a value of $m = 64$ was found to give good results. In the next step a quantification and standardisation of these coefficients is executed in order to speed up the calculation and minimise the memory requirements. Subsequently the wavelet coefficients are sorted according to the colour model, image channel, value and coordinates and are inserted into the leaf nodes of the index structure.

The determination of the n next neighbours of a query image is performed with the so called L^q metric [8] (modification of the Euclidian distance L^2), which considers deviations resulting from colour shifts, distortions and other errors due to the inaccurate query sketch. Let $Q[0,0]$ and $T[0,0]$ denote the coefficients of the scaling function, which depend on the average intensity of the colour channel. Let $\tilde{Q}[i,j]$ and $\tilde{T}[i,j]$ with $(0 \leq i,j \leq 127) \wedge \neg(i = j = 0)$ be the $[i,j]$-th k-largest and normalised wavelet coefficients. We simplify the calculation and set $\tilde{Q}[0,0] = \tilde{T}[0,0] = 0$, because these two values do not belong to any wavelet coefficient. The querying metric L^q is defined as:

$$L^q = ||Q, T||_q = w_{0,0}|Q[0,0] - T[0,0]| + \sum_{i,j} w_{i,j}|\tilde{Q}[i,j] - \tilde{T}[i,j]| \tag{15}$$

where Q is the query and T the target image. The values $w_{0,0}$ and $w_{i,j}$ are weights, which depend on the importance of the coefficients. They are empirically determined over a number of tests with manually selected images and grouped in classes marked with $W_1, \dots W_4$. For a detailed description of these weights and of the metrics the reader is referred to [3, 8].

After the wavelet decomposition of the query image Q the average intensity I_Q, the indexes and signs of the m largest coefficients for each colour channel c are determined. Subsequently we calculate the difference between the I_Q and I_T for all images T in the database. Finally for each of the wavelet coefficients $\tilde{Q}^c[i,j]$ we search all images, which have one m-largest coefficient with the same sign. The n next neighbours are the first n elements on the list sorted in descending order of the metric values $||Q,T||$.

5.1 Experimental results

The CLIMS methods for content based retrieval are tested on a database containing 1132 photographs of different categories. For the result evaluation we used the well-known measures Precision p and Recall r, defined as $p = |\mathcal{A} \cap \mathcal{B}|/|\mathcal{A}|$ and $r = |\mathcal{A} \cap \mathcal{B}|/|\mathcal{B}|$ respectively, where $\mathcal{A}$ is the desired and $\mathcal{B}$ the resulting part of the database. The interpretation of these values with respect to image retrieval is complicated, because the definition of $\mathcal{A}$ depends on the subjective user perception.

The retrieval output considers the 32 nearest pictures. The starting points for the test retrievals are 300×300 sketches of real images in the database, which are drawn by 10 different persons. There were no restrictions regarding the colours and draft style. But to prevent that too precise pictures were drafted, each test person had only 3 minutes to prepare the example sketch. Each query starts on the default combination of weights and colour models. If one combination failed in retrieving the correct picture, a next combination is selected. The weights $W_1, \ldots W_4$ are empirically compiled and tested. A detailed description can be found e.g. in [3, 8]. Our future work includes a development of automated methods for the adaptation of these values to the image classes in the database.

Table 1. Detailed results of the test scheme for the sketch retrieval

Weights and Hits in the first n pictures				
Colour / Weights	1-8	9-16	17-32	%
1. YUV/W4	32	5	6	43
2. YIQ/W4	14	9	2	25
3. HSL/W4	6	1	2	9
4. YIQ/W1	1	2	2	5
%	53	17	12	**82**

Table 1 summarises the results: approximately 82% of the sketches led to the desired target picture and only 18 remained without a direct hit. However, pictures similar to the target picture are returned in 14 of the 18 cases. If the sketches are considered as hits, a hit rate of 96% is achieved. If only static colour attributes are used for the retrieval, there are resulting values of approximately lower 30%. This method is acceptable only if pictures of a specific class, for example landscapes, are considered. In this case the calculation of precision and recall values is reasonable again and hit rates up to 68% can be achieved. Detailed representations of the performed measurements can be found in [7, 3]. Examples of results determined during a similarity search with the wavelet-based features are shown in the Figures 2 and 3.

Fig. 2. City-Query: An example image and the first 12 images of the retrieval.

6 Conclusions

CLIMS is a prototype for content based image retrieval in general image catalogues. It offers querying by example sketch and image and uses colour and wavelet based features for the comparison of images. All extracted features are stored in a relational database. A query starts with the application of the extraction algorithms on the given sample image or sketch. Thereby the search parameters are determined, which are subsequently used for the search of the corresponding index structures. These contain the a-priori calculated values for all images in the database. The result is a ranking of the n most similar images. The evaluation of the methods in a general image catalogue result in reasonable recognition rates.

Our future work includes the development and evaluation of other retrieval methods with a-priori and dynamic extracted features as well as the creation of suitable index structures. Furthermore we investigate different soft computing approaches for analysis and comparison of image content. Finally we develop a parallel cluster architecture for the efficient realisation of large image databases.

Fig. 3. Car-Query: The query sketch and the first 12 images of the retrieval.

References

1. S. Khoshafian, A. B. Baker. *MultiMedia and Imaging Databases.* Morgan Kaufmann Publishers, 1996.
2. J. Ashley et. al. Automatic and semi-automatic methods for image annotation and retrieval in QBIC. In *Proceedings of Storage and Retrieval for Image and Video Databases III*, pp 24–35, 1995.
3. I. la Tendresse. Wavelet features for content based image retrieval. Master's Thesis, TU Clausthal, 1999.
4. B. Momjian. *PostgreSQL: Introduction and Concepts.* Addison-Wesley, 2000.
5. M. Stricker, M. Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases III*, pp 381–392, 1995.
6. T. Chiueh. Content-based image indexing. In *Proceedings of the 20th VLDB Conference*, pages 582–593, Santiago, Chile, 1994.
7. C. Ruch. Content based search in image databases. Master's Thesis, TU Clausthal, 1997.
8. X. Wen, T.D. Huffmire, H.H. Hu, A. Finkelstein. Wavelet-based video indexing and querying. *Journal of Multimedia Systems*, 7:350-358, Springer, 1999
9. Y. Meyer Wavelets and Operators. In N.T.Peck, E.Berkson and J.Uhl, *Analysis at Urbana*, LNS 137, London Math. Society, 1989
10. I. Daubechies Orthonormal Bases of Compactly Supported Wavelets. *Communications on Pure and Applied Mathematics*, XL1, 1988
11. R. Coifman, G. Beylkin, V. Rokhlin Fast wavelet transforms and numerical algorithms I In *Communications on Pure and Applied Mathematics*, pp 44:141–183, 1991