AQUISAR: IMAGE RETRIEVAL IN UNDERWATER WEBCAM IMAGES

T. Kämpfe[1] *T. W. Nattkemper*[2] *H. Ritter*[1]

[1] Neuroinformatics Group, Faculty of Technology, University of Bielefeld, Germany
[2] Applied Neuroinformatics Group, Faculty of Technology, University of Bielefeld, Germany

ABSTRACT

This paper presents AQUISAR, a system for detecting interesting images shot by a webcam. Common webcams yield a huge amount of images, but most of them are considered to be not of interest for a lack of individual fascinating entities. Our system combines state-of-the-art techniques of CBIR and computer vision to detect images in the webcam image stream based on their contents. The content of an image is described by figures-background segmentation and statistical image segment features. Applying *query-by-example*, the database is searched for images showing contents similar to that reference example image. Since the webcam is located in an aquarium this work is settled in the upcoming field of underwater computer vision.

1. INTRODUCTION

Since in the Trojan Room faculty room at Oxford University the first webcam had been installed and the whole world could observe their coffee maker, the number of webcams has increased enormously. We will focus on the subset of webcams installed to observe natural scenes like animals in a reservation, e.g. [1]. Usually these cameras run round the clock or at least during daytime, whereas the interesting objects act just for a short period. Correspondingly, only a small subset of the images taken within one day are worth to be stored. Since it is not feasible for human users to observe the camera round the clock, a (semi-) automatic assistance is required to filter the interesting images out from the multitude of "boring" ones.

In webcam image retrieval nearly the same issues arise as the wide spread field of research content based image retrieval (CBIR) is dealing with. The forthcoming image set is quite large and a textual labeling by human users to represent the image content is not possible. Thus, statistical, low level features of the images are computed full-automatically and provide the ground basis for the filtering process. Nevertheless, the sole statistical features can not match the semantic interpretation a human user would prefer.

To bridge the gap between the user's interpretation and the low level feature domain, methods from image processing, pattern recognition, human-computer-interaction and relevance feedback are applied in recent proposed CBIR and multi media retrieval systems.

Webcams are usually installed for the purpose of observation of single places from a constant angle-of-view. Nevertheless a growing number of cameras switch between different angles-of-view and are located in quite misanthropic environments. Here we consider the extraordinary domain of underwater scenes. The multi-angle technique complicates an image segmentation and underwater images give rise to special issues caused by the physical attributes of water. In contrast to CBIR applications to standard image databases, color is a relatively improper image feature to describe underwater scenes since water absorbs most of the colors and the images are of low contrast and contain lots of noise. According to that underwater computer vision is a forthcoming field of research.

In the AQUISAR (**Aqu**arium **I**mage **S**egmentation **a**nd **R**etrieval) system we combine webcam image handling, content based image retrieval and underwater computer vision to propose the first multi-angle webcam CBIR system.

2. IMAGE RETRIEVAL IN AQUARIUM WEBCAM IMAGES

With the development of new photo and film technologies and the rising distribution of the internet the number of webcams is increasing rapidly. And although their utilization in science is quite reasonable, surprisingly little work for automatic extraction of interesting images exists. This may be caused by the highly diverse qualities of different webcams: Besides their common characteristic that each webcam provides a large set of images from a constant (set of) environment(s), there are less attributes in the images that are shared between image domains from different cameras.

For instance the camera installation is one major source of variability. Fixed camera positions occure as well as switching between a couple of determined orientations or

coninuous movings. Furthermore the update times are different. Consequently the suitability of common approaches to detect interesting image objects depends hardly on the regarded webcam. For example the proposed approaches used in the field of video surveillance can not be applied because of the low refreshing rate of the cameras. An additional feature of webcam images is that for the sake of transmission performance the amount of recorded data is limited and therefore low resolutions and color depths are used. Besides these technical characteristics the purpose of the webcams differs. Therefore the image domains vary as well as the categories of embodied entities.

Changing light conditions pose a further difficulty that arises in nearly every webcam image set: Most cameras are located in the open and even indoor light conditions change considerably during daytime. In this work we consider images from the London Aquarium [3] as a testbed for our system. The images spot a subregion of the aquarium, that contains sharks (sandtiger, brown sharks, zebra sharks), sting-rays and different sorts of fish swarms. The webcam switches between four settings and has an update interval of 5 seconds.

Underwater images depend on the special physical attributes of water: color extinction, reflection and scattering. The main features are the absence of color in larger depth, varying contrast, nonuniform and dim lighting and a lot of blur. Therefore, image retrieval suffers from the visibility conditions as well as from the characteristics of underwater objects.

3. THE AQUISAR SYSTEM

Our system AQUISAR (**Aqu**arium **I**mage **S**egmentation **a**nd **R**etrieval) performs the main steps necessary for retrieving interesting images in a set of images shot by the London Aquarium Webcam [3]. A sequence of preprocessing steps is implemented to calculate suitable image features. Afterwards these features are used to perform the retrieval of particular images.

3.1. Preprocessing of the Images

A fixed webcam takes pictures of predominantly a single scene with a quite unchanged background. In a set of images with the same background the image regions covered by changing entities can easily be detected via calculating difference images. The webcam in the London Aquarium switches between a couple of positions. For preserving the advantages of invariable backgrounds we group the N stored images $\mathbf{x}_i, i = 1, \ldots, N$ into clusters $C_j, j = 1, \ldots, N_{\text{pos}}$ according to the $N_{\text{pos}} = 4$ positions.

To obtain the desired partitioning, we employ *k-means*-clustering for feature vectors of the first 200 principal com-

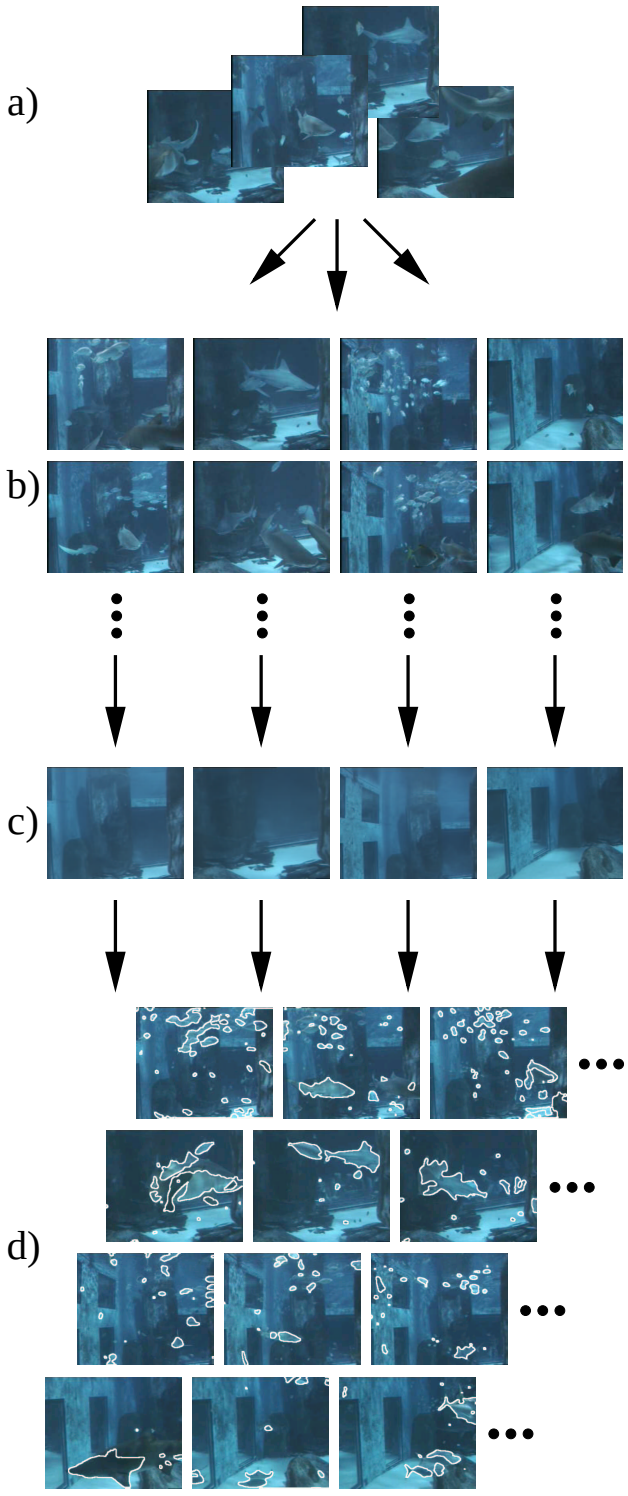

Fig. 1. Segmentation: A stored set of images (a) is grouped into four clusters with equal background (b). For each cluster the mean images provides prototypical views of an *empty aquarium* (c). Subsequently, an image segmentation is performed based on difference images (d).

ponents of the image autocorrelation matrix (see figure 1 (a+b)). In a next step a *region*-image $\mathbf{b}_i$ is computed, which assigns each pixel $\mathbf{x}_i^{xy}$ to a region $\mathbf{s}_{ki}$. To this end, a difference image $\tilde{\mathbf{x}}_i$ is computed first:

$$\tilde{\mathbf{x}}_i = |\mathbf{x}_i - \overline{\mathbf{x}}_j| \tag{1}$$

with $\mathbf{x}_i \in C_j$ and $\overline{\mathbf{x}}_j = \frac{1}{N_j} \sum_{\mathbf{x}_i \in C_j} \mathbf{x}_i$ is the average image of camera position j and N_j is the number of images taken from setting j. Note that the average images show an *empty* aquarium, as can be seen in figure 1c). From these difference images $\tilde{\mathbf{x}}_i$, label images $\mathbf{b}_i$ are computed which distinguish the background from possibly interesting coherent objects (i.e. fishes):

$$\mathbf{b}_i^{pq} = \begin{cases} k & , \text{ if } \tilde{\mathbf{x}}_i^{pq} \geq t \\ 0 & , \text{ otherwise} \end{cases} \tag{2}$$

where $\tilde{\mathbf{x}}_i^{pq}$ denotes the pixel value with the coordinates p, q of image $\tilde{\mathbf{x}}_i$ and t is a threshold calculated iteratively on the global grayvalue histogram [5]. The identifier k with $k \in [1, \ldots, K_i]$ is calculated in a preceeding step on the coherent binary objects that result from $\tilde{\mathbf{x}}_i^{pq} \geq t$ and is used to identify the various image regions:

$$\mathbf{s}_{ki} = \{\mathbf{x}_i^{pq} \mid \mathbf{b}_i^{pq} = k\} \tag{3}$$

K_i is the number of separate regions within image $\mathbf{x}_i$ and background pixel are labeled by $k = 0$.

For lack of specified features for underwater images like the aquarium images, we calculate a set of low-level image features for each region except the background. According to the special conditions in underwater computer vision, we assume that texture features are more suitable than color. Therefore we implemented two texture features (based on the fractal dimension resp. the co-occurence matrix [6]) and just one color feature (empirical mean and variance of HSV histograms).

3.2. Retrieval

The most intuitive and simple query to a webcam retrieval system is: "Show me interesting images!" Coming up with this task we have to answer two questions: (A) What is the meaning of interesting? (B) Which images achieve these specifications?

One obvious way to specify interesting images is presenting example images with a content the user consider absorbing. Based on this idea, we decided to use a *query by example*-framework in this initial analysis. This framework is suitable to detect images in a subject observation task, where an observer wants to know when a certain animal appears. With an example image containing the requested animal he can easily search for appropriate images.

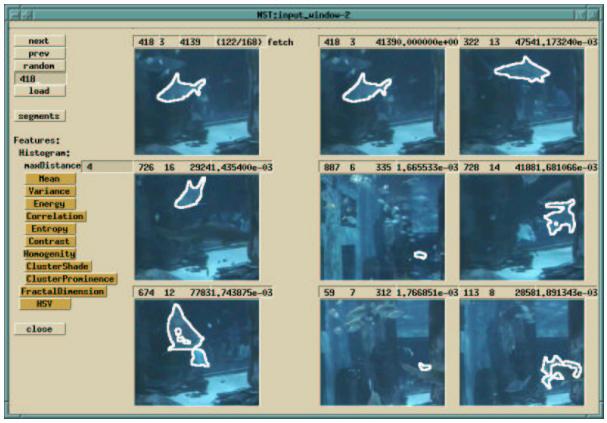

Fig. 2. Screenshot of the AQUISAR user interface presenting the query image (top left) and the eight result images. The crucial image regions are highlighted.

Depending on the quality of the segmentation result, the user may choose between various techniques to extract the query example $\mathbf{q}$: (i) choose an image region with a mouse click or (ii) pick up an explicit image region by enclosing the interesting image region by a sequence of mouse clicks.

To get the appropriate images the retrieval is performed as a similarity search. Therefore, the result is an ordered list $\mathbf{r}$ of the images or image regions:

$$\mathbf{r} = [\mathbf{s}_{ki}^{(1)}, \mathbf{s}_{ki}^{(2)}, \mathbf{s}_{ki}^{(3)}, \ldots] \tag{4}$$

with decreasing similarity values

$$d(\mathbf{q}, \mathbf{s}^{(u)}) \geq d(\mathbf{q}, \mathbf{s}^{(v)}) \quad \forall\, u, v \text{ with } u < v \tag{5}$$

$d(\cdot, \cdot)$ measures the similarity between two images. In this initial approach the Euclidean distance is calculated on the features $f_l(\mathbf{s}_{ki})$. Using the inverse of this distance we get the similarity measure. The first eight images of this list are presented in a graphical user interface (see figure 2).

4. OBSERVATIONS

The four positions of the London Aquarium webcam yield images consisting of quite similar image features. Since the clustering in this application was able to perfectly distinguish among these four views, we are very confident that image sets of other webcams with more differing positions can be clustered error free, too.

The mean of the images taken from the same camera setting renders a prototypical view of an empty aquarium (see figure 1c)). This can be regarded as the reference background to calculate difference images.

A quantitative evaluation of an unsupervised image segmentation result is quite difficult, in particular for an image

	mean	variance	min	max
AQUISAR	0.65	0.04	0.25	1
INDI	0.48	0.04	0.13	0.88

Table 1. Precision of retrieval results

set without a ground truth segmented data set. While the upcoming schemes to evaluate these issues often are based on color distributions we use in this initial phase a quantitative evaluation based on inspecting the segmentation results. In figure 1d) some images are presented, including the detected image regions. Although not every fish is cut out perfectly (especially very close and therefore big sharks are often detected just partially) the segments are suitable to calculate image features. An approximate border of an object is sufficient, since we use only color and texture features.

The most popular approaches to evaluate retrieval results are *precision-recall-diagrams* and derived measures [4]. But the absolute number of taken images containing a certain object is unknown, because an image set taken by a webcam is not labeled according to the contained objects. Therefore recall is not quantifiable. On the other hand, precision can easily be determined:

$$precision(i) = \frac{N_i^+}{i} \tag{6}$$

where N_i^+ is the number of interesting images (labeled by a human user) within the first i retrieved images.

In table 1 the achieved precision is presented. A human user has rated the results of 66 retrieval tasks and determined the N_i^+ for $i = 8$. For a comparison the same has been done with 58 tasks for the CBIR-system INDI [2].

The striking advantage of AQUISAR is that it can retrieve images with similar entities from *different* webcam settings, i. e. angles of view. The results of INDI contain just images taken from *the same* viewing angle as the reference image. This disadvantage of a conventional CBIR system like INDI is rooted in the fact that the main part of each aquarium image is covered by the background. Therefore the surrounding is dominant for calculating the result lists. Although INDI has some adaptive components to fit the users need based on relevance feedback, retrieving the same kind of fish taken from another position is not possible.

5. DISCUSSION

We have presented initial retrieval results of AQUISAR, a system currently under development for the task of identifying interesting images from an underwater webcam and supporting the user with browsing facilities, based on a query-by-example paradigm. We described several processing steps in AQUISAR specially tailored towards the special characteristics of underwater webcam images.

By comparing the precision of retrieval results with that of a previously developed, more general CBIR system we can show that taking the multi-angle nature of this image domain into account leads to a significantly improved retrieval accuracy. And the prospect for improvements according to the need of user can be in two directions: retrieval results and real-time usability. On the one side an advanced search strategy or similarity measure can enhance the retrieval results. Retrieval techniques used in CBIR systems, like relevance feedback or adaptable systems, can be built-in to enhance the detection of interesting images as well as the application of more advanced and domain specific image features, namely for determining the attribute "interesting".

On the other side the insight of detecting interesting images in a limited set of webcam images can be used to train an upgraded system, which detects interesting images online according to the web presentation.

Acknowledgement

We thank the "Morita Aquarium BV" (Trading as the the London Aquarium) for providing the webcam images used within this work.

This work is partially funded by the BMBF under contract 01IB 001B.

6. REFERENCES

[1] http://www.africam.com.

[2] T. Kämpfe, T. Käster, M. Pfeiffer, H. Ritter, and G. Sagerer. INDI – intelligent database navigation by interactive and intuitive content-based image retrieval. In *IEEE 2002 International Conference on Image Processing*, 2002.

[3] http://www.londonaquarium.co.uk/.

[4] H. Müller, W. Müller, D. McG. Squire, S. Marchand-Maillet, and T. Pun. Performance evaluation in content-based image retrieval: Overview and proposals. *Pattern Recognition Letters (Special Issue on Image and Video Indexing)*, pages 593–601, 2001.

[5] T. W. Ridler and S. Calvard. Picture thresholding using an iterative selection method. *IEEE Trans.*, SMC-8:630–632, Aug. 1978.

[6] M. Unser. Sum and difference histograms for texture classification. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, PAMI-8(1):118–125, 1986.