THE MOBILE MUSEUM GUIDE:
ARTWORK RECOGNITION WITH EIGENPAINTINGS AND SURF

Frederik Temmermans, Bart Jansen, Rudi Deklerck, Peter Schelkens and Jan Cornelis

Department of Electronics and Informatics, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel
IBBT, Gaston Crommenlaan 8, 9050 Gent

ABSTRACT

This paper investigates the performance of eigenpaintings, SURF and a combination of both methods in a query-by-picture system dedicated to artwork images. For this purpose a database containing photographs of 17 artworks was created. These photos were taken by visitors of the exhibition “The VUB Art Collection – On Display and in Your Hand” using the camera of a mobile phone. The resulting images present several challenging aspects, including reflections, different light conditions, variations in perspective viewing conditions etc. Experiments show that the method based on eigenpaintings classifies 85% of the artworks correctly and the method based on SURF 84%. The combined classifier shows an accuracy of 88%.

1. INTRODUCTION

The process of querying a database with an image as input is known as query-by-picture. It is a sub-domain of content based image retrieval [1], in which object recognition plays a crucial role. Object recognition often adopts features or configurations of features that reliably describe the same object in many images. Commonplace features are scale and orientation invariant properties of the edges and/or corners of the object. Tuytelaars et al. [2] provide an extensive overview of such invariant feature detectors, in which a distinction between global and local features is made. Global features, e.g. descriptors based on Principal Component Analysis (PCA) [3], grasp the overall composition of the image. Since they fail to discriminate between foreground and background, a pre-processing step is often required before global feature detection can take place. Local image features express properties of the object’s shape. The latter is often achieved by characterizing the corners, junctions or other regions with high curvature of the object. Technically, they are implemented as either corner detectors (e.g. [4]) or blob detectors (e.g. [5]), resulting in the well-known detectors “Scale Invariant Feature Transform” (SIFT) [6] and “Speeded Up Robust Features” (SURF) [7].

Advances with respect to camera resolution, processing, networking and display capabilities of smart phones, brought query-by-pictures applications to the mobile phone in a variety of domains, e.g. Paperboy (newspapers), Kooaba, Snooth (wine), Google Goggles, USnap (advertisement) and many more. It is also an application domain that currently is actively studied by standardization committees, e.g by the Visual Search and JPSearch groups of respectively the working groups WG11 (MPEG) and WG1 (JPEG) of the ISO/IEC JTC1/SC29 standardization committee.

This paper explores the use of the query-by-picture technology on mobile phones for artwork recognition in museums. By taking a picture of an artwork, the visitor receives additional information on his mobile phone. Compared to static displays next to the artwork, this allows for personalized information and for multimedia items to be delivered on an individual basis. Image recognition is performed using a method based on global features (a principal component based approach dubbed eigenpaintings) and a method based on local features (SURF). It is also explored whether both approaches are indeed complementary and improve recognition accuracy.
This paper is organized as follows: Section 2 gives an overview of the image preprocessing steps and the classification methods applied after preprocessing. Section 3 provides experimental results. Finally, Section 4 discusses the obtained results and the strengths and weaknesses of our approach.

2. IMAGE PROCESSING STEPS

2.1. Foreground extraction
To ensure that recognition is performed on the artwork and not on the background of the image, the first step in the processing chain consists of extracting the artwork, i.e. the foreground. This paper adopts a multi-step approach for this extraction task. First a mean shift filter smoothens the image. Thereafter, the Shi-Tomasi [8] corner detector identifies the candidate corners in the image. Next, the bounding rectangle around all candidate corners is extracted. The final bounding quadrangle is then defined by taking the four candidate corners, closest to the four corners of the initial bounding rectangle. This way, perspective deformations are considered, but with a preference for quasi rectangle bounding quadrangles. Based on this bounding quadrangle, a perspective transform is applied to obtain a rectified foreground image.

This ad hoc approach for artwork segmentation relies on the fact that the artwork is presented on a uniform white background and that the biggest region sufficiently contrasting to the background is the artwork. Although successful in most of the presented artworks, this method does not generalize to other problems.

2.2. Artwork recognition based on eigenpaintings
For computational reasons the images obtained from the artwork extraction (2.1) are downsampled to 256-by-256 pixels and converted to grayscale. From all images in the database a mean image is computed. The eigenvectors (referred to as eigenpaintings) of the covariance matrix of the set of mean subtracted images of the artworks are computed. Each of the training images is then represented as a weighted sum over these eigenpaintings [3]. Also the query image is projected on these eigenpaintings to obtain a weight vector. In this first prototype, the Euclidean distance between the weight vector of the query image and the weight vector of a training image is calculated to assess the similarity between both images. For any query image, the recognition result is the training image with highest similarity to the query image. The query strategy will be refined in future research.

2.3. Artwork recognition based on SURF
SURF is a scale- and rotation-invariant interest point detector and descriptor [7] and is a successor of SIFT [6]. The similarity score between two images is based on the number of matches between their interest points. For each interest point in the first image, the matching interest point in the second image is the interest point with a descriptor vector depicting the smallest Euclidean distance. A match is considered ambiguous if the Euclidean distance between the two matching interest points is smaller than 0.65 times the distance between the interest point and the second closest match. These ambiguous matches are removed from the set of matching interest points. This is a well known nearest neighbor matching strategy (e.g. [7]).

2.4. Combination of eigenpaintings and SURF
Given the fact that eigenpaintings are global features and SURF detects local features, improved accuracy could be obtained from combining both approaches. In a hybrid classifier, for any query image the similarity scores to each of the training images are calculated independently using both SURF and eigenpaintings. The similarity scores for the query to each of the training instances are normalized, such that they sum to one, for both algorithms. The similarity score between any two images in the hybrid classifier is then a weighted sum over the similarity scores of the two classifiers. Although the weights can be learnt based on training images, we opted in the initial experiments for
equal weights for both classifiers. Based on this similarity score, the best match is returned from the database of training images.

3. EXPERIMENT

In the autumn of 2009, the exhibition entitled “The VUB Art Collection – On Display and in Your Hand”, exposed artworks belonging to the university. The collection includes works by Sam Dillemans, Jan Burssens, Liliane Vertessen, Marcel Marieën, Nele Tas and Koen Broucke. 17 of these works are included in the classification task investigated in this paper. Visitors took images using an iPhone 3G with a 2MP camera (1200 by 1600 pixels), provided by the organizers. The captured images were sent to a server for processing and compared to a database of prototype images for each of the 17 classes. Based on the recognized artwork, multimedia information was streamed to the user.

The experiments reported in this paper are performed offline, on a database of 284 images captured at the exhibition, as explained above. For each of the 17 classes, the database contains between 8 and 22 images. In our experiment, each artwork (class) is represented by 4 random selected prototype images. The remaining images are used as test images. The distance between any query image and any class is defined as the sum of the distances between the artwork and each of the four prototypes. Under these conditions, the method based on eigenpaintings classifies 184 of the 216 images correctly (85% overall or 84% class-per-class average), SURF classifies 181 images correctly (84% overall or 87% class-per-class average) and the hybrid classifier classifies 191 images correctly (88% overall or 91% class-per-class average).

Normalized confusion matrices based on eigenpaintings, SURF and the hybrid classifier are visualized in the tables shown in Figures 3, 4 and 5 respectively. These tables reveal that all images of class 15 are misclassified using the eigenpaintings technique, while SURF classifies all these images correctly. This can be explained by the failure of the background extraction for this particular work. The eigenpaintings technique requires a good registration of the content. SURF on the other hand, only needs to find enough matching points, which is feasible in this case because of the high frequency content in this work.

Looking at class 5, we see that all test images of this class are misclassified by SURF, while eigenpaintings classifies 63% images correctly. In this case, SURF often does not find any matches with any of the pictures in the training set. This is because artwork 5 shows a rather uniform white region and a rather uniform dark region (see Figure 2). The white region is often recognized as background during preprocessing and SURF fails to identify features in the black region. Because the mistake in background also occurs for some of the images in the training set, the eigenpaintings technique still scores fairly well.

In both cases explained above, the hybrid classifier scores better than the failing classifier. Although it never scores better than the best performing technique, in average, it scores better than both techniques individually.

Additional experiments showed that using color information in the eigenpaintings rather than intensity information does not influence the obtained results.

Figure 3: Normalized confusion matrix based on eigenpaintings. An entry at row $i$, column $j$ gives the percentage of test images of class $i$, classified as class $j$.

Figure 4: Normalized confusion matrix based on SURF. Note that in this case it is possible that rows do not sum to 1. This is because in some cases it is possible that SURF does not return a match.

Figure 5: Normalized confusion matrix based on the combined classifier.
Figure 6: Two images of the same artwork (nr 2). The images clearly illustrate differences in reflections, perspective viewing and light conditions.

4. DISCUSSION

The reported experimental results confirm that query-by-picture on a database of artworks is feasible, using eigenpaintings or SURF. SURF and eigenpaintings showed very similar performance. Our preliminary results show that the combination of both approaches into a hybrid classifier results in better performance. It is to be investigated whether these results hold with larger test sets, with more image classes and under different weighting schemes between both classifiers. Also, the performance of the “Mobile Museum Guide” was only evaluated using pictures of artwork and it is unclear how it generalizes to other domains.

In the approach proposed in this paper, a preprocessing step extracts the foreground from the images and recognition is only performed on the foreground. Without including this step, matching based on properties in the background cannot be avoided. Additionally to the experiments described above, we’ve explored image recognition on the entire images, without foreground extraction. In that case the accuracy of eigenpaintings drops to 50%, while the accuracy of SURF remains unaffected.

Compared to other domains (e.g. recognition of buildings in city landscapes), the recognition of artwork might be a simpler problem. Nevertheless, inspection of the 17 training images shown in figure 1 shows that lighting conditions are very different among the images and that many images are influenced by light reflections in the glass covering some artworks (e.g. 2, 7, 8, 14-17, see Figure 6). Also, the artwork segmentation is problematic for those artworks having large uniform white regions without a strong boundary (e.g. 5 and 9) or for the pictures including part of the floor (e.g. 13) or the ceiling (e.g. 15). The three classifiers are robust to these deformations.

5. ACKNOWLEDGEMENT

This work was performed in context of the iBrussels project (www.ibrussels.org). A project funded by the Brussels Capital Region and supported by the Brussels Regional Informatics Centre. Peter Schelkens holds a postdoctoral fellowship with the Fund for Scientific Research (FWO). This work is part of the “Digital Painting Analysis” project, supported by the Apple ARTS program.

6. REFERENCES