Image-Based Retrieval and Identification of Ancient Coins

Martin Kampel, Vienna University of Technology
Reinhold Huber-Mörk, Austrian Research Centers
Maia Zaharieva, University of Vienna

The illegal trading and theft of coins appear to be a major part of the unlawful antiques market, so methods for computer-aided identification of ancient coins can help conserve cultural heritage. Traditionally, people have identified ancient coins by manually searching coin publications and auction catalogues; today, large digital collections are available. For coin classification, textual descriptions provide relevant information. Object identification, on the other hand, relies on unique features that distinguish a given object from all other members of the same class, including variations in the hammering process, die, mint signs, shape, scratches, wearing, and so on.

Content-based image retrieval (CBIR) is an emerging technology in the field of cultural heritage, relying on textual and pictorial descriptions of coin collections to aid classification and identification. CBIR is commonly used to group similar images on the basis of features such as color, texture, and shape. In general, this approach depends on the variety of content: the more distinguishable the objects are (for example, holiday snapshots), the better the recognition. For particular areas of computer vision, such as object identification and object class recognition, retrieval methods are specifically designed to exploit class properties and background knowledge.

In a typical CBIR setting, the user is interested in all images similar to a single query; we, on the other hand, don’t use a typical CBIR setting because our goal is to identify one specific coin in a collection of coin images. (See the sidebar “Related Work on Coin Identification” for more details.)

Challenges and Approaches
Figure 1 illustrates the challenges of coin identification. The Fitzwilliam Museum in Cambridge, England, which contains more than 40,000 digital coin descriptions, provided us with an image database comprising 2,400 images of 240 different ancient coins of the same class. Independent numismatists selected the coins. They recorded both sides of each coin five times using different tech-
Techniques: three scans in different rotations with an offset of 120° and two photos under varied illumination conditions of each coin side. The scans provide an image resolution of up to 700 × 700 pixels. The images acquired by camera are of lower quality—some are blurred and some are at most 350 × 350 pixels. Each row shows the same coin acquired by different devices and varying in terms of condition and orientation.

At first sight, all the coins bear the same characteristics. However, different dies produced the coins shown in each row. This data set is ideal for thoroughly testing identification methods because all the coins are similar—all issued in the time of, or at least in the name of, Alexander the Great. Some of the coins are from much later and were minted in places around the Black Sea, in Egypt, in modern-day Turkey, and in Iran. All the coins follow the same basic standard: the face shows a head of Heracles in a lion skin, and the back shows the god Zeus, seated on a throne facing left. Nevertheless, the huge range of detail in the coins’ minor variations enables experts to deduce each coin’s mint and date.

Our approach relies on combining shape and local descriptors to capture a coin’s unique shape characteristics and die information, respectively. For identifying ancient coins, we propose fusing information using a pre-selection step based on shape matching (comparing objects characterized by their shape representations), combined with a subsequent identification step based on coin die matching. The processing of coin images involves segmentation, shape feature extraction, and local descriptor extraction.

Knowledge-Based Segmentation

The separation of an object of interest from the background is commonly called segmentation. Because of textured backgrounds, the presence of other objects in the image, and non-homogeneous or poor illumination, straightforward low-contrast methods based on global thresholding of image intensity tend to fail. The use of background knowledge about the image content, in our case coins, largely helps to reduce segmentation errors.

In situations where explicit knowledge of objects’ properties is available, this knowledge can be used to steer segmentation parameters. For example, researchers have used the compactness measure to find an intensity threshold in images showing circular spot welds. Similarly, researchers have localized ancient coins by thresholding the local-intensity range (defined as the difference between maximum and minimum gray levels in a local window).

Typically, the shape of modern coins is close to a circle, whereas ancient coins deviate more from this shape. So, we employ a measure of compactness $c_t$ related to a threshold $t$ defined as

$$ c_t = \frac{4\pi A_t}{P_t^2}, $$

where $A_t$ is the area of the region covered by a coin and $P_t$ is the coin’s perimeter. $A_t$ and $P_t$ are obtained by connected-components analysis applied to the binary image, which is derived from thresholding the intensity-range image. Figure 2a shows an intensity image of a coin, Figure 2b is the corresponding intensity-range image, and Figures 2c–2g show thresholded images for different selections of $t$ along with calculated values for $c_t$. Figure 2h shows the image thresholded...
The fact that ancient coins are generally not perfectly circular facilitates identification. Moreover, it’s often possible to infer a coin image’s spatial orientation from its shape. We’re especially interested in a coin’s shape described by its edge, which is usually expressed as a set of pixel positions sampled along the edge. From a numismatic point of view, a coin’s shape serves as a first clue in the process of coin identification and discrimination.

Methods for image-based recognition of modern coins include artificial neural networks; edge features; gradient directions; eigenspaces; and color, shape, and wavelet features. Researchers have successfully used a mechatronic device to sort more than 300 tons of modern coins. To support the evaluation of recognition methods for modern coins, the EC-sponsored Network of Excellence program made available its Coin Image Seibersdorf (CIS) benchmark data set. Researchers recently presented additional approaches for classifying ancient coins.


Several approaches, including object and texture recognition, robot localization, symmetry detection, wide baseline stereo matching, and object class recognition, apply local descriptors in computer vision. In spite of their success and generality, these approaches are limited by the distinctiveness of the features and the difficulty of appropriate matching. (For a thorough survey and evaluation of the performance of local features and their repeatability in the presence of rotation, scale, illumination, blur, and viewpoint changes, see the work by Krystian Mikolajczyk and Cordelia Schmid. Recently, SIFT (scale-invariant feature transform) features were applied to recognizing ancient coins. The goal of information fusion is to achieve better decisions by combining sources than what using a single source can achieve. For modern coins, Reinhold Huber has suggested a Bayesian fusion method that combines classification hypotheses obtained from obverse and reverse sides of coins.

### References


Related Work in Coin Identification

at the optimal level $t_{opt}$ with highest compactness.

**Deviation from Circular Shape Matching**

To match ancient coin shapes, researchers recently applied a shape description algorithm and a robust correlation algorithm. On the basis of concepts in the latter approach—describing the shape border as a deviation from a circle—we developed a method called deviation from circular shape matching (DCSM). Because DCSM uses background knowledge of coin shapes (which are close to being circular), it isn’t well suited for matching general shapes but performs quite well on ancient coins.

DCSM proceeds as follows. On the basis of the segmentation of the ob-
ject, we trace the border of the coin in the thresholded image and obtain a list of border pixels. This list is resampled to \( l \) samples using equidistantly spaced intervals with respect to the arc length. Figure 3a shows this operation. Figure 3b shows the polar distances between sampled points and the circle as a vector for the entire border of the coin.

We obtain a 1D descriptor—that is, a curve describing the border—by fitting the coin edge to a circle and unrolling the polar distances between sample points and the fitted circle into a vector. We derive the center \( s_c = (x_c, y_c) \) of the fitted circle from the center of gravity, and the radius \( r \) is the mean distance between the center and all sample points \( s_i = (x_i, y_i) \). The 1D representation is given by \( D = (d_1, \ldots, d_l) \), where

\[
d_i = \left( \|s_i - s_c\| - r \right) / r, \quad r = 1, \ldots, l.
\]  

(2)

Dividing by \( r \) makes the representation invariant with respect to scale.

Figure 2. (a) Segmentation of a coin (b) from thresholding the local-intensity range, (c–g) for different selections of a threshold \( t \) and compactness \( c_t \), and (h) at the optimal level. The coin is detected at the threshold level achieving highest compactness.

(a) Intensity image
(b) Intensity-range image
(c) Binary image: \( t = 5, c_t = 0.096 \)
(d) Binary image: \( t = 25, c_t = 0.868 \)
(e) Binary image: \( t = 45, c_t = 0.879 \)
(f) Binary image: \( t = 65, c_t = 0.859 \)
(g) Binary image: \( t = 85, c_t = 0.879 \)
(h) Binary image: \( t_{opt} = 49, c_t = 0.888 \)

Figure 3. Deriving a coin’s shape description from the fitted circle. (a) We use deviation of sampling points along the coin border from circular shape and (b) arrange the polar distances between sampled points and circle into a vector.
We then compare two shape descriptions (SDs) corresponding to coins by means of a linear combination of global and local shape matching. We derive the local matching from the difference of Fourier shape descriptors, whereas the correlation coefficient between the curves serves as the global shape matching.

We use the mean squared distance (MSD) between the magnitude values of the Fourier coefficients as a local measure of dissimilarity. That is, the MSD between two coins A and B is

$$\text{MSD} = \sum_{i=1, \ldots, l-u} \| \text{sd}_A(i) - \text{sd}_B(i) \|_p,$$

where $$\| \cdot \|_p \in \{1, 2\}$$ is the $$L_p$$ norm. The lower ($$v \geq 1$$) and upper ($$u \geq 0$$) offsets for the Fourier descriptors are small constants that are used to limit errors stemming from imprecise circle fitting and quantization noise.

We obtain a global shape matching from the normalized cross-correlation (NCC) coefficient $$\text{ncc}(u)$$ for a shift of $$u$$ samples. The maximum $$\text{NCC} = \max_{i=1, \ldots, l} \text{ncc}(i)$$ is used as a measure of global shape match. The position of the maximum is related to the rotation angle between the compared coins.

The overall measure of shape dissimilarity becomes

$$\text{DISS}_{AB} = \alpha \text{MSD} + (1 - \alpha) \frac{1 - \text{NCC}}{2},$$

where the weighting factor $$\alpha \in [0, 1]$$ controls the influence of local and global dissimilarity terms.

**Local Descriptors**

A wide variety of local features exists in the literature. Different descriptors emphasize different image properties such as intensity, edges, or texture. We focus on the two descriptors that show outstanding performance with respect to changes in illumination, scale, rotation, and blur.

The *scale-invariant feature transform* (SIFT) descriptor is based on gradient distribution in salient regions. At each feature location, we select an orientation by determining the peak of the histogram of local image gradient orientations. Subpixel image location, scale, and orientation are associated with each SIFT feature vector (4 × 4 location grid × 8 gradient orientations). We identify interest points at peaks (local maxima and minima) of Gaussian function applied in scale space, and we eliminate all keypoints with low contrast or keypoints that are localized at edges using a Laplacian function.

*Speeded-up robust features* (SURF) are fast scale- and rotation-invariant features. The descriptor captures distributions of Haar wavelet responses within the neighborhood of an interest point. Each feature descriptor has only 64 dimensions, which results in fast computation and comparison (4 × 4 location grid × 4 wavelet responses in horizontal and vertical directions).

Figure 4 visualizes selected descriptors (a set of 16 histograms aligned in a 4 × 4 grid, each with 8 orientation bins) on the front and back sides of an ancient coin.

We match the local descriptors by identifying the first two nearest neighbors in Euclidean space. We accept a descriptor only if the distance ratio to the second-nearest neighbor is below a given threshold (for example, 0.8). However, our experiments showed that for the case of lower interclass dif-
ferences (all classes are coins), a lower distance ratio tends to keep more distinctive descriptors while eliminating a great part of the false matches. For our coin recognition application, we determined a distance ratio of 0.5 experimentally to be most adequate. Figure 5 shows an example of matching interest points between two images of the same coin using the SIFT descriptor.

**System Workflow**

The workflow of our system is illustrated in Figure 6. Taking a digital image as input, the first step (image segmentation) is to detect the image region showing the coin. In the next step (feature extraction), the system uses the segmented region to obtain relevant features from the coin image. To decrease computational costs and increase the overall recognition rate, the system preselects possible coins. Finally, the coin recognition step generates an ordered list of coin classes and coin IDs along with a reliability grade for each class and ID.

Figure 7 shows the graphical front end for shape matching and preselection. The query result for the image is shown on the left. The list of filenames in the center shows the dissimilarity for each image. The matching coin, shown on the right, is rotated using the estimated rotation angle.

Figure 8 shows the graphical interface for stand-alone die matching and combined shape and die matching. In this interface, the user can select a coin image to be matched on the top-left part of the window, and matching methods and combinations of methods from the lower-left part of the window. The right part of the window shows found matches and some information about the matching (for example, the number of descriptors and determined matches).

For the semantic encoding, we chose the CIDOC-CRM ontology. We use the CIDOC-CRM classes and properties not only to encode the descriptions provided by the coin recognition process and by numismatists but also to capture and encode the relationships among events, people, places, objects, and time spans directly referred to by the sources. The ontology is flexible enough to allow a perfect conceptualization of the meaning contained in each sentence and to encode them in a machine-understandable set of documents to be used in different semantic applications. To map different data schemata, the data-mapping tool AMA (Archive Mapper for Archaeology) offers a base template for coins. AMA aims at facilitating the mapping of existing data structures to a standard one. This digital-reference collection of abstract coin types stores items with the features (textual descriptions from numismatists together with descriptors from the coin recognition process) that make the coin shown in Figure 8 suitable as an optimal representative of its class.
Results
Experimental results characterize the subcomponents’ performance. We evaluated the subcomponents on a set of representative test data, described earlier.

Shape matching
We achieved the best performance of DCSM using 256 sample points, a weighting parameter $\alpha$ of 0.025, and lower and upper offsets ($u$ and $v$) of 3 each. We also used leave-$n$-out accuracy estimators, which are the average nearest-neighbor classification scores obtained from matching each of the 2,400 images to all other images. The leave-$5k$-out, where $k$ is the number of classes, refers to training and test sets of equal sizes. The leave-$9k$-out result corresponds to training sets containing single images per class. The data show that even in the case of a single example per class in the training set, the accuracy stays above 90 percent. The result of 99 percent for the leave-one-out estimator means that the system incorrectly classified only 24 images. A detailed look at the incorrectly classified images revealed that they are characterized by a rough border, which was caused by the discretization of the imaging process and disturbed the Fourier space description.

Die matching
SIFT features are 128 dimensional vectors that describe the neighborhood of interest points. The high dimensionality of the SIFT descriptor and the large number of interest points detected is a drawback of SIFT at the matching step. On the contrary, SURF features are efficiently computed using integral images and use 64 dimensional vectors. For the data set of ancient coins we were studying, the number of points detected with SURF was reduced by 70 percent on average. However, the reduction of information captured leads to loss of valuable information.

Figure 9 shows the identification rate of both descriptors for different sizes of training sets. In every iteration of the experiment, we used $N$ random images per coin as the training set. Because each side of a coin is pictured five times, $N = 1, ..., 4$ coins are used for training, and the method is evaluated on the remaining ones. The results shown in the figure are averaged rates over 10 iterations for each $N$. We determined the number of iterations experimentally; further iterations didn’t have any significant impact on the results.

The results show that SIFT clearly outperforms SURF independent of the size of the training set and matching strategy. However, the quality and size of the training set definitely influence identification performance.

Combined approach
We combine shape and local descriptors to increase the identification rate. Preselection based on shape matching lets us restrict the required comparisons for matching local features. As a result, we speed up the identification process and achieve better accuracy. Because our shape descriptor doesn’t vary with mirroring, preselection can...
be performed on either the whole available coin data set (that is, the pre-selected set can contain images of the second side of the coin) or just the relevant side. Figure 10 summarizes the results derived from leave-one-out estimation for different preselection sets obtained by DCSM and based on the whole available coin set. The figure also shows the identification accuracies with respect to different acquisition devices.

In contrast to the results we previously presented, the identification rates of both SIFT and SURF features lie close to each other. This is due to the shape descriptor’s level of accuracy. Preselection of up to 10 image candidates discards a major part of the mismatches of local features. At this level, SURF becomes a competitive approach again, because it has a short computational time and significantly high accuracy rate.

We found the following identification rates for the single and combined descriptors:

- the DCSM shape descriptor—90.23 percent,
- the local descriptor’s SIFT features—71.77 percent,
- the local descriptor’s SURF features—56.24 percent,
- the DCSM/SIFT combined descriptor—94.11 percent, and
- the DCSM/SURF combined descriptor—95.16 percent.

We performed shape-based preselection according to the side of the test coin shown. We found that the identification rate using DCSM is lower than when we used leave–out estimation. Because DCSM is designed to match coins regardless of which side is presented, the accuracy obviously drops first when considering matching of coin sides. Combining both descriptors clearly holds the top-performing accuracy rate of 95.16 percent for correctly identified coins.

Finally, the system is integrated into a Web application that acts as a portal. It integrates multiple coin archives and provides powerful search capabilities. Figure 11 shows the final Web interface. A native XML database is used to store and index the coins’ XML data, which is structured according to the defined ontology.

Acknowledgments

We thank Mark Blackburn of the Fitzwilliam Museum for providing test images and numismatic expertise and Achille Felicetti of the University of Florence for input concerning data management. This work was partly supported by the European Union under grant FP6-SSP5-044450. This article reflects only the authors’ views; the European Community is not liable for any use that might be made of the information contained herein.
The Authors

Martin Kampel is an assistant professor of computer vision in the Pattern Recognition and Image Processing Group at the Vienna University of Technology. His research interests include 3D vision and cultural-heritage applications, visual surveillance, and image sequence analysis. Kampel received his PhD in computer science with distinction from the Vienna University of Technology. Kampel is a member of the International Association for Pattern Recognition and the IEEE. Contact him at kampel@prip.tuwien.ac.at or www.prip.tuwien.ac.at.

Reinhold Huber-Mörk is a researcher in the Smart Systems Division’s High-Performance Image Processing business unit at the Austrian Research Centers. His research interests include image processing and pattern recognition for machine vision applications. Huber-Mörk received his PhD in computer science with distinction from the University of Salzburg. Contact him at reinhold.huber@arcs.ac.at or www.smart-systems.at.

Maia Zaharieva is a doctoral candidate in the Interactive Media Systems Group at the Vienna University of Technology. Her research interests include image processing, pattern recognition, and video retrieval. Zaharieva received her MSc in business informatics from the University of Vienna. Contact her at zaharieva@ims.tuwien.ac.at or www.ims.tuwien.ac.at.

References


For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.