

Habilitationsschrift

**Recognition and Reconstruction in
Archaeology and Surveillance**

eingereicht an der Technischen Universität Wien
Fakultät für Informatik

Dipl. Ing. Dr. techn. Martin Kappel
Favoritenstr. 89/10, 1100 Wien

Wien im Oktober 2008

Abstract

This cumulative habilitation thesis gives a short introduction to the area of my research in the last years as a member of the Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group at the Vienna University of Technology. It is about Computer Vision in general, and it is about 3D data acquisition and processing to achieve 3D reconstructions of real objects and about developing specific techniques for scene analysis in particular. The first part gives an introduction to related research. In a first thematic approach a workflow from 3D data acquisition towards 3D volume reconstruction is given. The second thematic issue briefly covers object recognition, whereas the third topic is on the consecutive steps from motion detection in image sequences until scene interpretation and its evaluation. Next, a short summary of a collection of seven representative publications of my work carried out in these areas is given, followed by a discussion of how they contribute to the further advancement of research. Moreover the original contributions of each paper and of the cumulative thesis are presented. The second part consists of the seven papers themselves, where the first four papers present 3D Vision techniques. The next two are concerned with local features for object recognition and colorimetric features for color constancy respectively. The final paper deals with a novel approach to represent chrominance information more suitable for robust background modelling and shadow suppression.

Acknowledgement

This thesis would not have been possible without the support of numerous people at PRIP (Pattern Recognition and Image Processing) group at the Vienna University of Technology. I particularly wish to thank:

- **Walter Kropatsch** for supporting me as a researcher and teacher at our university. In particular his trust in me and the freedom he gave me were one of the most important contributions to my career.
- **Robert Sablatnig** for his valuable suggestions and the excellent team work especially regarding our teaching responsibilities and joint research. He has supported me in many organizational matters and in the integration in our faculty.
- **Allan Hanbury** for fruitful discussions and for carefully proof-reading this thesis.

Furthermore I would like to thank all my former and current colleagues for their teamwork and good sense of humor providing an excellent working atmosphere, special thanks to *Maia Zaharieva*, *Philipp Blauensteiner*, *Horst Wildenauer*, and *Sebastian Zambanini*.

I am grateful to several foundations and institutions for direct and indirect financial support of my studies: the EU under the FP5 and FP6 projects 3D MURALE (IST-1999-20273), COINS (FP6-SSP5-044450), AVI-TRACK (AST-CT-502818), the Austrian Science Foundation (FWF) under grant P13385 and the company Ing. Johann Schiessel Elektronik + Computertechnik in the project Computer Aided Bank Surveillance (CABS).

Last but not least, my deepest gratitude goes to my wife **Brigitta** for her constant support and to my children **Miriam** and **Felix**, who are showing me that there is life and reason beyond work.

Contents

1	Overview	1
1.1	3D object reconstruction	3
1.2	Object recognition	4
1.3	Image sequence analysis	6
1.4	Summary of papers	8
1.5	Discussion of papers	14
	Bibliography	20
2	Octree-based Fusion of Shape from Silhouette and Shape from Structured Light	34
3	An Automated Pottery Archival and Reconstruction System	39
4	On 3D Mosaicing of Rotationally Symmetric Ceramic Fragments	50
5	Rule based System for Archaeological Pottery Classification	55
6	Recognizing Ancient Coins based on Local Features	64
7	Color Classification of Archaeological Fragments	77
8	Improved Motion Segmentation based on Shadow Detection	82

Chapter 1

Overview

Humans receive information about the world to a large extent through their visual sense. Retinal sensors convey stimuli to the visual cortex, where the visual perception is processed [36]. The first computers powerful enough to carry out image processing tasks appeared in 1960. Since then machines have been built to see, which lead to a wide range of techniques in the area of computer vision. Nowadays computer vision is not solely a scientific topic anymore, arts, hobbyists, especially a growing audience from those interested in multimedia are increasingly evident.

In [104] L.G. Shapiro defines the goal of computer vision as to make useful decisions about real physical objects and scenes based on sensed images. Besides the quality of the algorithms, the overall effectiveness of a computer vision system depends critically on acquired sensory data [40].

The topics in this thesis can be subdivided by the sensed data they rely on (see Figure 1 for an illustration): 3D images represent 3D objects, which allow us to consider a 3D volume as part of the entire 3D world. A 3D representation is a transition of an object in the real 3D world to an object-centered co-ordinate system, allowing the object descriptions to be viewer independent [107]. A digital image is defined as a 2D image $I[r, c]$ represented by a discrete 2D array of intensities [104]. A sequence of images represents how images change due to motion of objects, camera, or light source [103].

This habilitation thesis contributes to three aspects of the analysis and processing of digitally sensed images:

- 3D object reconstruction: acquisition and processing of 3D data for 3D

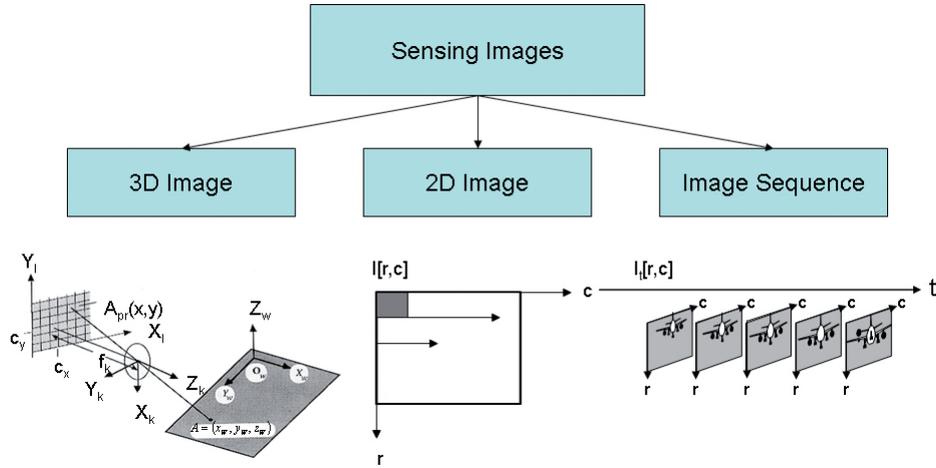


Figure 1: Sensing Images

reconstructions of real objects.

- Object recognition: investigation of local image descriptors and color constancy for object identification
- Image sequence analysis: shadow suppression for robust moving object detection.

This habilitation thesis is a collection of different papers, each describing a recognition or reconstruction task for a specific type of application. The areas of application are archaeology and visual surveillance. Let us first have a closer look into 3D vision (Section 1.1) and then analyse different features necessary for object recognition (Section 1.2). In Section 1.3 we describe the consecutive steps from motion detection in image sequences until scene interpretation. These last three sections provide a condense overview of the research areas. The focus was put on those areas and topics which are relevant to understand the papers in this thesis in a wider context. Section 1.4 summaries the papers, followed by a discussion on how they contribute to the further advancement of research. Chapters 2 to 8 consist of the collected papers themselves.

1.1 3D object reconstruction

In order to quantitatively describe the relationship between 2D image structures and their corresponding real world structures, methods for extracting 3D information have to be investigated. According to Marr [86] we see 3D vision as a 3D object reconstruction task to describe 3D shape in a co-ordinate system independent of the viewer. The shape of a 3D object is represented by a 3D model. Two main classes of 3D models are identified [107]: volumetric models represent the “inside” of a 3D object explicitly, while surface models use only object surfaces. Unlike 3D models, depth maps or range images describe relative distances from the viewer of surfaces detected in the scene [86]. For 3D object reconstruction the following three major steps, as illustrated in Figure 2, are identified: (1) data acquisition, (2) range image processing and (3) model reconstruction.

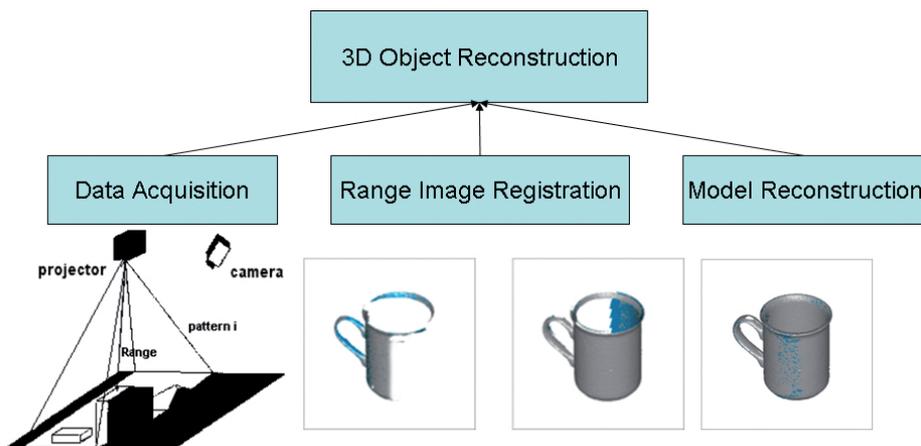


Figure 2: 3D Object Reconstruction

Data acquisition is the first and one of the most important tasks in a chain of 3D reconstruction tasks, because the data quality of the sensed images influences the quality of the final results [9]. El. Hakim specifies in [26] the quality of data by a number of requirements: high geometric accuracy, capturing all details, photo-realism, full automation, low cost, portability, flexibility in applications, and efficiency in model size. The techniques that aim to recover shape from intensity images are described by *Shape from X* [107], in [11] F. Blais reviews the last 20 years of range sensor development.

Each view acquired by a range sensor represents a portion of the object. Multiple views of one the same object have to be combined to produce a complete model of the object. The process of combining range data into one single coordinate system is called **range image registration** [18]. The approximate transformation between the views can be obtained either automatically by a calibrated robot or shifting unit or by the computation of the rigid transformation that maps the data from one view into that from another view. The most commonly used algorithm for registering is the Iterative Closest Point (ICP) algorithm [10]. ICP iteratively improves the registration of two overlapping surfaces by calculating the unique transformation that minimizes the mean square distances of the correspondences between the two surfaces.

The final step **model reconstruction** refers to a surface fusion process, which consists of multiple tasks like reducing large data sets, outliers removal, surface triangulation, hole filling, etc. The goals are to minimize a global registration error, caused by the propagation of noise from one surface patch to the other and to preserve the topology of the original object.

With an increasing capability for rapidly acquiring large quantities of 3D data, many existing algorithms face issues of computational complexity and data representation, and need to be revisited. Hot topics in 3D Vision i.a. deal with scene capture and self calibration [93, 5], 3D modeling under time-varying motion, deformation, or illumination [20], 3D modeling from large and complex data sets [95], or 3D view registration and recognition [100].

1.2 Object recognition

The ability to recognize objects in images forms an essential area of research in computer vision. The primary goal of object recognition is the detection and localization of arbitrary objects in an image [94]. Recognition generally requires an exact object description in a form suitable for a classifier. A primary difference between various approaches is the type of features used for recognizing objects, see [110] for an overview. Based on the papers selected for this habilitation thesis we briefly report on three of them:

Shape for example is a carefully investigated object property [107], which e.g. is described by global shape descriptors, geometric invariants, or edge points. Problems are caused by improper object segmentation, object occlusions, fragmented object boundaries or varying camera view-points. A comparison of shape descriptors based on curvature-scale space, wavelets, visual parts, Zernike moments, multilayer eigenvectors and directed acyclic graphs is given in [74]. Recently, further advances were reported using a hierarchical likelihood cut-off scheme [88], using the shape context approach [8] to compare shapes based on the earth movers distance [78], and using hierarchical deformable shapes, the so-called shape tree [28].

The application of **local features** in computer vision is manifold ranging from object [30] and texture recognition [75] to robot localization [91], symmetry detection [79], wide baseline stereo matching [117], and object class recognition [89]. In spite of their success and generality, these approaches are limited by the distinctiveness of the features and the difficulty of appropriate matching [30]. For a survey and evaluation on the performance of local features in the context of their repeatability in the presence of rotation, scale, illumination, blur and viewpoint changes refer to [90].

Color based object recognition belongs to appearance based methods, which traditionally extract global features from the whole image [110]. A common approach is the use of color histograms to compare color distributions of the image [25]. If the light reflected by an object is measured and used for object recognition, the object's color drift due to illumination changes has to be investigated [33]. Possible solutions are to apply a color constancy algorithm [25] or define a measure that is invariant to changes of the illuminant [37].

Current approaches are often tailored towards solving specific problems, like recognizing pedestrians [76] or face recognition [73]. General solutions need features that are invariant to image scaling, translation, and rotation, and invariant to illumination changes and projective transformation [110]. Current state of the art e.g. addresses the problem of learning object models from cluttered images [29], or training from a view images [116].

1.3 Image sequence analysis

A sequence of images represents how images change due to motion of objects, camera, or light source [103]. Figure 3 illustrates three steps necessary for analysing an image sequence for scene interpretation. The first step **motion detection** attempts to locate connected regions of pixels that represent the moving objects within the scene; there are many ways to achieve this including frame to frame differencing, background subtraction and motion analysis (e.g. optical flow) techniques. Background subtraction methods [108, 39, 122] store an estimate of the static scene, which can be accumulated over a period of observation; this background model is subsequently applied to find foreground (i.e. moving) regions that do not match the static scene. In Figure 3 (top image) black pixels indicate (moving) foreground and white pixels indicate (static) background.

The subsequent process moving **object tracking** takes as input the result from the motion detection stage and commonly applies trajectory or appearance analysis to predict, associate and update previously observed objects in the current time step. Tracking algorithms have to deal with motion detection errors and complex object interactions e.g. merging, occlusion, fragmentation, non-rigid motion, etc. For example the Kanade-Lucas-Tomasi (KLT) feature tracker [105] combines a local feature selection criterion with feature-based matching in adjacent frames; this method has the advantage that objects can be tracked through partial occlusion when only a sub-set of the features are visible. To improve the computational efficiency of the tracker in [112] motion segmentation is not performed globally to detect the objects. Instead, the features are used in conjunction with a rule based approach to find correspondences between connected foreground regions; in this way the KLT tracker *simultaneously* solves the problems of data association and tracking without assumption of a global motion for each object. For a detailed overview on visual surveillance of object motion and a taxonomy for object tracking methods see [38]. Figure 3 (center image) shows tracked objects by their bounding boxes together with their trajectories.

The goal of an object recognition process for **scene interpretation** is to identify at least the object category and at most the object category, size

and precise spatial attributes (e.g. orientation, centroid etc). Model based methods (e.g. Sullivan [109]) can be applied to locate the objects of interest in the scene. An alternate approach is to train a classifier to distinguish the different object types (e.g. Collins *et al* [21]); a major drawback with this approach is the scalability to classifying multiple objects from multiple cameras, especially when there are minor differences between some object types or when objects appear vastly different under perspective transformations. In Figure 3 (lower image) an alert together with a scene description is shown.

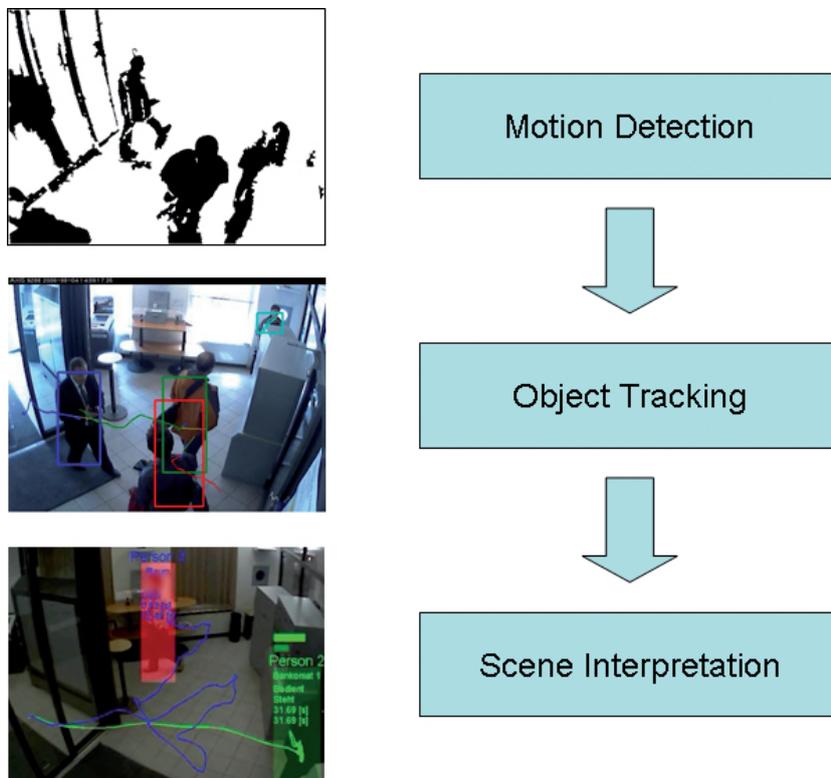


Figure 3: Image sequence analysis: workflow

Data fusion combines the tracking data measured by the individual cameras to maximise the useful information content of the observed scene. In [113] visual surveillance aspects of a distributed system focusing on a discrete nearest neighbour Kalman filter approach for data fusion are presented.

Hot topics in the area of image sequence analysis are i.a. Multi-camera and Multi-modal Sensor Fusion [17], scene understanding and event recog-

dition [14] or practical issues like performance over time, robustness against lighting conditions and shadows [1]. In order to assess reliability and robustness of existing approaches and to facilitate their comparability there is also an emerging demand for quantitative evaluation of segmentation and tracking quality [27]. Especially in the case of outdoor surveillance, where illumination changes, weather conditions, shadows, and occlusions strongly impact the motion segmentation quality and consequently all subsequent steps, robustness is of crucial importance.

1.4 Summary of papers

The papers in this collection have been chosen to be included in this thesis as most representative of my work conducted in the area of computer vision. Three main areas can be identified: 3D data processing with particular attention to 3D acquisition and 3D processing, 2D image analysis focusing on recognition and color, and image sequence analysis. All of them are either journal or IEEE proceedings publications, except for one published by Springer Lecture Notes in Computer Science (LNCS). As a first author of all of these papers my contribution to each of these articles and the scientific work behind is substantial. In general, computer vision research can be seen as a collaborative effort, where cooperations with colleagues and students are a prerequisite for any successful development. Therefore none of the papers in this thesis is a single-author paper. The contribution of this cumulative thesis is a collection of different papers, each describing a recognition or reconstruction task for a specific type of application. Four papers deal with 3D vision topics, two papers are concerned with 2D images analysis focusing on object recognition and color constancy. These contributions represent applied research in the area of cultural heritage. The remaining paper covers image sequence analysis with respect to motion detection as a contribution to visual surveillance.

The collection starts with a paper [67] on combining two acquisition approaches for 3D volume reconstruction. The original contribution is a novel strategy for 3D data acquisition of arbitrarily shaped objects. The approach overcomes the drawbacks of structured light scanning devices in the case of

occlusions. This work was initiated and supervised by the author, while the implementation was carried out by diploma student Srdan Tosovic under the guidance of the author together with Robert Sablatnig.

It is followed by two papers [57, 61] on processing 3D data focusing on 3D reconstruction and 3D mosaicing. A novel, fully automated pipeline for 3D modelling from acquired range images is presented, considering the different approaches and analysing all the steps involved. The method avoids any manual intervention, which is usually needed for any 3D reconstruction process. Another contribution is a model based approach for 3D fragment assembly, which avoids the complex and time consuming task of finding matching fragments of a broken object in order to reconstruct the original object. These two papers [57, 61] and the next one [62] with Robert Sablatnig are based on work largely done during his supervision of my PhD thesis.

The next paper [62] deals with curvature analysis for 3D object classification. The novel shape based scheme presented offers a typological classification into discrete types and allows a systematic analysis of large quantities of acquired material.

One paper [70] is on object recognition based on local image features. My original contribution to the topic focuses on an assessment of the quality and efficiency of local features for recognizing of very similar objects. The results lead to a reliable strategy for the identification and classification of these objects. This work was initiated and supervised by the author, while the implementation was carried out by PhD student Maia Zaharieva.

This is followed by one paper [51] dealing with a color constancy technique. A novel approach for the estimation of colorimetric features used for color calibration is shown. The method proposed allows accurate color estimation, which plays an important role in any subsequent color based classification process. For this paper Robert Sablatnig provided his expert knowledge on the archaeological framework and context.

The collection ends with recent work [68] on adapting colour models for shadow detection in image sequences. My contribution to the topic focuses on a novel approach for shadow suppression. The results lead to a more robust background model, allowing better performance of an object tracking or scene interpretation process. Allan Hanbury provided his expert knowledge on

the Hue, Luminance and Saturation (IHLS) color space, whereas Philipp Blauensteiner and Horst Wildenauer contributed to the development of the system and provided the annotated dataset for the experiments.

List of papers

This thesis contains the following papers:

- Martin Kampel, Srđan Tosovic, and Robert Sablatnig. Octree-based Fusion of Shape from Silhouette and Shape from Structured Light, *Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission (3DPVT02)*, IEEE Computer Society, pp. 754-757, 2002. [67]
- Martin Kampel and Robert Sablatnig. An Automated Pottery Archival and Reconstruction System, *Journal of Visualization and Computer Animation*, John Wiley & Sons, Ltd., Vol. 14(3), pp. 111-120, 2003. [57]
- Martin Kampel and Robert Sablatnig. On 3D Mosaicing of Rotationally Symmetric Ceramic Fragments, *Proceedings of 17th International Conference on Pattern Recognition*, IEEE Computer Society, Vol. 2, pp. 265-268, 2004. [61]
- Martin Kampel and Robert Sablatnig. Rule based System for archaeological Pottery Classification, *Pattern Recognition Letters*, Elsevier Science Inc., 28(6):740-747, 2007. [62]
- Martin Kampel and Maia Zaharieva. Recognizing Ancient Coins based on Local Features, *Advances in Visual Computing*, Springer. LNCS 5358, pp.11-22, 2008. [70]
- Martin Kampel and Robert Sablatnig. Color Classification of Archaeological Fragments, *Proc. of 15th International Conference on Pattern Recognition*, IEEE Computer Society, Vol. 4, pp. 771-774, 2000. [51]
- Martin Kampel, Horst Wildenauer, Philipp Blauensteiner and Allan Hanbury, Improved Motion Segmentation based on Shadow Detection, *Electronic Letters on Computer Vision and Image Analysis*, Vol. 6(3), pp. 1-12, 2007. [68]

A brief summary of each of the papers is presented next:

- **Fusion:** In [67] a novel algorithm for the automatic reconstruction of a 3D model using two different 3D algorithms is presented. To acquire the 3D shape of objects with concave surfaces is complicated, since occlusions of the objects surface are introduced and can only be resolved by taking multiple views. Therefore, the 3D reconstruction is based on a sequence of images of the object taken from different viewpoints with different algorithms: shape from silhouette and shape from structured light. The output of both algorithms are then used to construct a single 3D model. It is shown that combining different acquisition methods leads to a solution for the automatic creation of models of arbitrarily shaped objects.
- **Reconstruction:** The system presented in [57] uses the profile of a rotationally symmetric object, which is the cross-section of the object in the direction of the rotational axis of symmetry, to classify and reconstruct it virtually. The input data for the estimation of the profile is a set of points produced by a 3D acquisition system. By registering a front and a back view of the object the profile is computed and measurements like diameter, area percentage of the complete object, height and width are derived automatically. We show an automated way for complete 3D reconstruction of a fragmented object, based on a novel method for range image registration.
- **Mosaicing:** The work in [61] aims to develop a solution for automated reassembly of broken artifacts. Generally the 3D reconstruction of arbitrary objects from their fragments can be regarded as a 3D puzzle. In order to solve it we identified the following main tasks: 3D data acquisition, orientation of the object, classification of the object and reconstruction. Matching candidates are identified by a pre-classification tasks, final matches of two neighboring artifacts are based on Euclidean distances. The pros and cons of the approach are shown by practical examples on real 3D objects.
- **Classification:** The paper [62] deals with a shape based classification scheme based on primitives. In order to segment a profile sec-

tion into shape primitives, rules based on human expert knowledge are created. The input data for the estimation of the profile is a set of points produced by the acquisition system. A function fitting this set is constructed and later on processed to find the characteristic points necessary to classify the original object. We show an automated way for 3D object classification, which outperforms manual approaches by its speed, accuracy and reliability.

- **Recognition:** In [70] we present an extension and combination of local image descriptors relevant for recognition of rather unvarying objects. Interest points are detected and their appearance is described by local descriptors. Object identification is based on the selection of similar images based on feature matching. We evaluate the performance of various interest point detectors and local descriptors with respect to recognition. Besides an automated way for recognizing ancient coins we show that *SIFT* combined with a *Difference of Gaussian Detector* outperforms other techniques in terms of the recognition rate.
- **Color:** In this work [51], a novel approach for the estimation of colorimetric features for color constancy is proposed. The technique exploits the fact that the spectral reflectance of certain materials varies slowly in the visible. We explain how the acquisition system is calibrated in order to get accurate color information. We present a simple and therefore practical approach for allowing color based object classification.
- **Motion Detection:** The paper [68] discusses common colour models for background subtraction and problems related to their utilisation. A novel approach to represent chrominance information more suitable for robust background modelling and shadow suppression is proposed. Our method relies on the ability to represent colours in terms of a 3D-polar coordinate system having saturation independent of the brightness function; specifically, we build upon an Improved Hue, Luminance, and Saturation space (IHLS). In the proposed framework we show an improved motion detection strategy based on robust shadow suppression.

I have also contributed to other peer reviewed scientific papers. Data acquisition strategies (2D and 3D) are investigated in [54, 127, 46, 45, 53, 99, 51, 52, 56]. Details on the Fusion of 3D acquisition approaches can be found in [66, 67, 41] and specific tasks in the area of 3D data processing are described in [96, 23, 50, 49, 58, 65, 47, 24, 98, 59, 60, 64, 2, 84, 63, 55, 83, 48]. Object recognition based on image descriptors are further investigated in [128, 125, 123, 124, 69, 83, 42].

Contributions to surveillance with respect to motion detection can be found in [119, 12], regarding object tracking in multi camera environment can be found in [112, 31, 115, 3, 114, 72] and regarding evaluation strategies in [43, 111, 4, 118].

1.5 Discussion of papers

The papers that I selected for this habilitation thesis are related to each other in that they all deal with *specific aspects* of image data processing embedded in *real world* applications. These aspects are **3D information extraction** for the reconstruction of real objects, **object recognition** based on shape and image features, and **color analysis** towards color constancy and color modeling.

Computer vision research in the application domain exists since the early research stages: on one hand the results achieved in basic research are converted into real world constraints, in order to prove that the methods work and are of practical use. On the other hand application oriented research may originate from a goal driven development process: the *relative* importance of observations and expectations depends on the goals [7], which may be defined by the application. Applied computer vision, better known as machine vision is increasingly seen as a key component for maximising quality, productivity and efficiency in many automated manufacturing and process control tasks [106].

In addition to now well established industrial applications, many new and challenging application areas continue to be explored and developed across a wide range of other disciplines. Examples include innovative applications evolving in nano-technology, document analysis, robotics and reverse engi-

neering, surveillance and cultural heritage. For the latter several problem solutions are described in this habilitation thesis.

Cultural heritage is at a point where it can benefit greatly from the application of computer vision methods, and in turn provides a large number of new, challenging and interesting conceptual problems and data for computer vision [87]. Thus, the research work is growing into a new field, like the cultural heritage field on the example of archaeology. For example the studies on archaeological fragments have led to the FP5 EU project 3D MURALE [97] and more recently the work on ancient coins is funded by the FP6 EU project COINS [126]. This trend is also shown by special workshops on cultural heritage applications at major vision conferences like the *International Conference on Computer Vision (CVPR)*, the *International Conference on 3D Imaging and Modeling (3DIM)* or the *European Signal Processing Conference (EUSIPCO)*. Furthermore call for papers from journals like *IEEE Intelligent Systems, AI and Cultural Heritage*, or *IEEE signal processing magazine, Special Issue on Visual Cultural Heritage* show a high interest from the computer vision society.

Over the last decade, research in the field of visual surveillance has gained a considerable amount of attention: i.a. a recent increase in the number of surveillance cameras has led to a strong demand for automatic methods of processing their outputs. Unlike the previous area of application visual surveillance is an application of computer vision with strong industrial interests and thus financial support. In our case the studies on image sequences have led to the FP6 EU project AVITRACK [12], and to projects funded by industrial partners (e.g. Center Systems GmbH, Erste Bank, ...) or by the Austrian Research Promotion Agency (FFG). All major computer vision conferences have special tracks and sessions dedicated to algorithms for detection, classification and tracking of moving objects and therefore reflect a high interest from the research community.

The *specific aspects* previously mentioned are treated separately below:

3D information extraction

The first three papers (Sections 2 to 4) deal with 3D information extraction. The paper in Section 2 “**Octree based fusion of shape**

from silhouette and shape from structured light” deals with a combination of two different 3D acquisition techniques in order to create a volume model of a real object. Original aim was to find a solution to overcome the drawbacks of 3D acquisition techniques based on structured light (e.g. occlusion handling, high number of views necessary for complex objects). A major obstacle to the wider use of 3D object reconstruction and modeling is the extent of manual intervention needed. Such interventions are currently massive and exist throughout every phase of a 3D reconstruction project: collection of images, image management, establishment of sensor position and image orientation, extracting the geometric detail describing an object, and merging geometric, texture and semantic data. This paper proposes a novel approach for the automatic and fast creation of 3D volume models of arbitrarily shaped objects. The idea behind this work is that a surface description is acceptable only if it is consistent with all the images captured of it. The approach contributes to archaeological research by providing the exact volume of arbitrary vessels in an automated and accurate way. The technique is not restricted to a specific application, it can provide models of any kind. The paper and previous versions of it motivated other researchers like the group of H. Rushmeier from Yale University to extend the idea, e.g. combining shape from silhouettes with shape from shadows [101]. In [82] the technique has been used as a starting point for measuring the shape of transparent objects.

Whereas the paper just discussed focuses on 3D acquisition techniques the second paper (see Section 3) **“An automated pottery archival and reconstruction system”** deals with the processing of range data for 3D surface reconstruction of a single object. The paper tackles the problem of range image registration with non overlapping views. Our solution to this problem is to apply a model based registration technique, as shown for rotationally symmetric objects. Besides that the paper demonstrates a complete workflow for virtual reconstruction based on the estimation of a profile section. From the application point of view the approach presented provides an automated tool for fast and accurate processing of large number of archaeological objects

and therefore contributes to overcome the main drawbacks of conventional archaeological documentation. An important assumption for the proposed methods is that the shape of the objects is assumed to be rotationally symmetric. The impact of the paper is twofold: researchers from the two different disciplines, computer vision and archaeology, integrated the paper into their research: this work together with previous versions are cited by archaeologists like V. Nautiyal and J. Clark [92] or A. Karasik et al. [34], and by computer vision researchers like Cooper et al. [121], or H.C.G. Leitão [77]. J.A. Barceló lists the paper in [6] when summarizing most prominent work on shape analysis of archaeological objects. The range image registration technique has also been used as a starting point for work on the inspection of gear box teeth [13]. Furthermore our contributions to the cultural heritage research domain over the last years have been acknowledged by the community, and will bring the two main conferences the *International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage* (VSMM) in 2009 and the *International Symposium on Virtual Reality, Archaeology and Cultural Heritage* (VAST) in 2010 to Vienna. (I have the pleasure to be Co-Chair of both events).

The next paper (Section 4) **“On 3D mosaicing of rotationally symmetric ceramic fragments”** focuses on 3D reconstruction with respect to fragment assembly. Unlike the previous paper, where the 3D reconstruction of a single object was described, this paper presents a fragment matching technique for various fragments of one and the same object. The preselection of matching candidate fragments is performed by a classification based on geometric features and its properties (see Section 5). Like the previous paper the method proposed works for rotationally symmetric objects only. The fragment matching relies on a pre-alignment through the computed axis of rotation (as described in 3) and a resulting two-degrees-of-freedom search space. Matching pairs are identified by minimizing the Euclidean distance between corresponding points. The paper fits into various attempts for solving 3D jigsaw puzzles [32, 15, 35], and is unique in its pre-classification step. It contributes to archaeological research by providing both, a real and

virtual, complete reconstruction of the original object. Solving automatic jigsaw puzzles is a computer vision problem, that tackles various topics like shape description, boundary matching, pattern recognition, feature extraction and matching. Other researchers like P. Cignoni [19] investigate the capabilities of such a reconstruction techniques from a practical point of view, whereas M. Callieri et al. [16] described it with respect to graphics especially visualization. C. Maiza et al. [81] made use of the ideas in their work on matching between a sherd and a shape model.

Object recognition

The next two papers in this habilitation thesis (Sections 5 and 6) concern recognition tasks. They differ from those described in the previous paragraph in that they present recent work. The paper “**Rule based system for archaeological pottery classification**” describes a shape based scheme for identifying local shape contours. It is based on hierarchical segmentation of the longest profile line, which is selected as a representative of the object. The profile is represented by spline functions, in order to use shape parameters for classification and recognition [55]. The evaluation is based on the processing of a limited number of real 3D objects. The resulting classification into specific types of objects is of utmost relevance for the archaeological documentation process. Karasik and Smilanski [71] extended the basic idea and introduced the concept of a “mean profile”, in order to be less affected by local, non-representative deformations. In contrast to Gilboa’s strategy [34], where the classification is based on hand drawn profiles, this paper presents a fully automated classification approach for processing 3D Data acquired by a 3D scanner. Our method is not restricted to archaeological objects, e.g. other applications based on partially localized shape contours like the representation of facial features [80] are reported. On the other hand the rules are defined for specific objects, and therefore not generalizable.

The paper “**Recognizing ancient coins based on local features**” in Section 6 tackles the problem of reliable object identification in the

process of recognition and traceability of stolen cultural heritage. Existing approaches for object recognition are investigated for their effectiveness for use with ancient coins. What makes this application special and challenging for object recognition, is that all the coins are very similar. The poor evaluation results of state-of-the-art techniques led us to propose a novel approach based on a combination of shape and local descriptors to capture the unique characteristics of the coin shape and die information respectively. In [44] we additionally present results on a data set of 2400 coin images of 240 different coins. A scientific requirement in numismatic research is to assign a coin its correct number according to a reference book. The approach presented provides such an assignment automatically. Independently of the applied local descriptor a combination of shape and local features achieved a top performing identification rate of 95.16%. This revised and extended version of the paper is currently under review for journal publication [44]. The work on ancient coin recognition has led to the organization of the *1st International COIN Recognition Workshop* in August 08, where experts from both numismatics and computer vision met (I had the pleasure to be the General Chair).

Color analysis

The final two papers of this habilitation thesis (Sections 7 and 8) concentrate on color calibration and color models for background subtraction. The first of these (Section 7) “**Color classification of archaeological fragments**” describes a novel approach to get colorimetric accurate information. This is important, since different illuminations, cameras or acquisition setups produce different color output for one the same object. Our approach exploits the fact that spectral reflectance of certain groups of objects like fruits or archaeological objects vary slowly in the visible. Archaeological research benefits from the method presented by getting accurate color information of archaeological objects, which usually varies depending on the user and light conditions. Besides inspiring other researchers (e.g. [85, 22, 120]) the work had

already a great practical impact. A dutch company¹ specialised in highly automated agricultural production systems integrated the idea into an industrial process dealing with automated fruit and vegetable grading. Fruit and vegetables are classified into quality classes based on their color appearance, color constancy consequently is of utmost importance.

The final paper “**Improved motion segmentation based on shadow detection**” proposes a novel approach for representing chrominance information more suitable for robust background modelling and represents very recent work. The ability to effectively segment moving objects under a wide range of disturbing conditions is a critical requirement for all subsequent tasks for event detection. If the quality of motion segmentation deteriorates (e.g. moving objects are not detected or segmented poorly), subsequent tracking or categorisation will yield erroneous results or simply fail. We show that an Improved Hue, Luminance and Saturation (IHLS) color space outperforms other color spaces for background modelling. The additional contribution of the approach is that we deal with the problem of unstable hue values at low saturation by modelling the hue-saturation relationship using saturation-weighted hue statistics. The effectiveness of the proposed method is shown in an experimental comparison with approaches based on RGB, Normalised RGB and HSV. Setiawan et. al. [102] apply the idea and propose another variant of a Gaussian Mixture Model by exploiting the (IHLS) color space. The proposed shadow detection technique is integrated in a visual surveillance system brought to market by a french airport infrastructure supplier.

¹www.ellips.nl

Bibliography

- [1] D. Abe, E. Segawa, O. Nakayama, M. Shiohara, S. Sasaki, N. Sugano, and H. Kanno. Robust small-object detection for outdoor wide-area surveillance. *IEICE Transactions on Information and Systems*, E91-D(7):1922–1928, July 2008.
- [2] K. Adler, M. Kampel, R. Kastler, M. Penz, R. Sablatnig, and K. Hlavackova-Schindler. Computer Aided Classification of Ceramics - Achievements and Problems. In W. Börner and L. Dollhofer, editors, *Proc. of 6th International Workshop on Archaeology and Computers*, pages 3–12, Vienna, Austria, 2001.
- [3] J. Aguilera, D. Thirde, M. Kampel, M. Borg, G. Fernandez, and J. Ferryman. Visual surveillance and monitoring for an airport’s apron. In O. Chum and V. Franc, editors, *Proc. of the 11th CVWWS6: Computer Winter Vision Workshop*, pages 5–10, 2006.
- [4] J. Aguilera, H. Wildenauer, M. Kampel, M. Borg, D. Thirde, and J. Ferryman. Evaluation of motion segmentation quality for aircraft activity surveillance. In T. Tan R. Chellappa, J. Ferryman, editor, *Proc. of the Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 293 – 300, 2005.
- [5] D.G. Aliaga and Y. Xu. Photogeometric structured light: A self-calibrating and multi-viewpoint framework for accurate 3d modeling. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR08)*, pages 1–8, 2008.
- [6] J. A. Barcelo. *Computational Intelligence in Archaeology*. IGI Global, Information Science Reference, Henshey (VA), USA, July 2008.
- [7] H. G. Barrow and J. M. Tenenbaum. Representation and use of knowledge in vision. *ACM SIGART Bulletin*, 52:2–8, 1975.
- [8] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

- [9] J. A. Beraldin, C. Atzeni C., G. Guidi, M. Pieraccini, and S. Lazzari S. Establishing a Digital 3d Imaging Laboratory for Heritage Applications: First Trials. In *Proc. of the Italy-Canada 2001 Workshop on 3D Digital Imaging and Modeling Applications*, Padova, 2001. on CD-ROM.
- [10] P. J. Besl and N. D. McKay. A method for registration of 3d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
- [11] F. Blais. A Review of 20 Years of Range Sensor Development. *Journal of Electronic Imaging*, 13(1):231–240, 2004.
- [12] P. Blauensteiner and M. Kampel. Visual surveillance of an airport’s apron - an overview of the avitrack project. In *Digital Imaging in Media and Education, Proc. of the 28th Workshop of the Austrian Association for Pattern Recognition (OAGM/AAPR)*, pages 213 – 220, 2004.
- [13] S. Boukebbab, H. Bouchenitfa, H. Boughouas, and J. M. Linares. Applied iterative closest point algorithm to automated inspection of gear box tooth. *Computers and Industrial Engineering*, 52(1):162–173, 2007.
- [14] F. Bremond. *Scene Understanding: perception, multi-sensor fusion, spatio-temporal reasoning and activity recognition*. HDR Universit de Nice-Sophia Antipolis, 2007. Habilitation thesis.
- [15] G. C. Burdea and H. J. Wolfson. Solving Jigsaw Puzzles by a Robot. *IEEE Transactions on Robotics and Automation*, 5(6):752–764, 1989.
- [16] M. Callieri, P. Cignoni, F. Ganovelli, G. Impoco, C. Montani, P. Pingi, F. Ponchio, and R. Scopigno. Visualization and 3d data processing in the david restoration. *Computer Graphics and Applications*, 24(2):16–21, 2004.
- [17] A. Cavallaro. Multi-sensor object detection and tracking (editorial). *Signal, Image and Video Processing*, 1(2):99–100, 2007.
- [18] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image Vision Computing*, 10(3):145–155, 1992.
- [19] P. Cignoni and R. Scopigno. Sampled 3d models for ch applications: an enabling medium or a technological exercise? *Journal on Computing and Cultural Heritage (JOCCH)*, 1(1), 2008.
- [20] B. Cihan and Y. Yemez. Time varying surface reconstruction from multiview video. In *IEEE International Conference on Shape Modeling and Applications*, pages 47–51, 2008.
- [21] R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson. A system for video surveillance and monitoring. In *Tech. Report CMU-RI-TR-00-12*, May 2002.

- [22] D.B. Cooper, X. Orriols, S. Velipasalar, E.L. Vote, M.S. Joukowsky, B.B. Kimia, D.H. Laidlaw, D. Mumford, A. Willis, S. Andrews, et al. Assembling virtual pots from 3D measurements of their fragments. *Proc. of the 2001 Conference on Virtual Reality, Archaeology, and Cultural Heritage*, pages 241–254, 2001.
- [23] J. Cosmas, T. Itegaki, D. Green, N. Joseph, L. Van Gool, A. Zalesny, D. Vanrintel, F. Leberl, K. Schindler, K. Karner, M. Gervautz, S. Hynst, M. Waelkens, K. Pollefeys, T. Vereenoghe, R. Sablatnig, M. Kampel, and P. Axell. Providing multimedia tools for recording, reconstruction, visualization and database storage/access of archaeological excavations. In *4th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage (VAST03)*, 2003.
- [24] M. Drenik and M. Kampel. An evaluation of low cost scanning versus industrial 3d scanning devices. In *IEEE International Conference on Image and Signal Processing (CISP 2008)*, volume 3, pages 756 – 769. IEEE Computer Society, 2008.
- [25] M. Ebner. *Color Constancy*. Wiley, first edition edition, 2007.
- [26] S. F. El-Hakim, J. A. Beraldin, and M. Picard. Detailed 3D Reconstruction of Monuments using multiple Techniques. In W. Boehler, editor, *Proc. of ISPRS-CIPA Workshop on Scanning for Cultural Heritage Recording*, pages 13–18, Corfu, 2002.
- [27] T. Ellis. Performance Metrics and Methods for Tracking in Surveillance. In *IEEE Workshop on Performance Evaluation of Tracking and Surveillance, Copenhagen, Denmark*, pages 26–31, June 2002.
- [28] P. F. Felzenszwalb and J. D. Schwartz. Hierarchical matching of deformable shapes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR07)*, pages 1–8, 2007.
- [29] V. Ferrari, F. Jurie, and C. Schmid. Accurate object detection with deformable shape models learnt from images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR07)*, pages 1–8, 2007.
- [30] V. Ferrari, T. Tuytelaars, and L. V. Gool. Simultaneous object recognition and segmentation from single or multiple model views. *International Journal of Computer Vision*, 67(2):159–188, 2006.
- [31] J. Ferryman, M. Borg, D. Thirde, F. Fusier, V. Valentin, F. Bremond, M. Thonnat, J. Aguilera, and M. Kampel. Automated scene understanding for airport aprons. In *Proc. of the Australian Joint Conference on Artificial Intelligence*, pages 593 – 603, 2005.

- [32] H. Freeman and L. Garder. Apictorial jigsaw puzzles: The computer solution of a problem in pattern recognition. *IEEE Transactions on Electronic Computers*, 13(2):118–127, April 1964.
- [33] J.M. Geusebroek, R. Boomgaard, A. Smeulders, and T. Gevers. Color constancy from physical principles. *Pattern Recognition Letters*, 24(11):1653–1662, 2003.
- [34] A. Gilboa, A. Karasik, I. Sharon, and U. Smilansky. Towards computerized typology and classification of ceramics. *Journal of Archaeological Science*, 31:681–694, 2004.
- [35] D. Goldberg, Ch. Malon, and M. Bern. A global approach to automatic solution of jigsaw puzzles. In *SCG '02: Proc. of the 18th annual symposium on Computational Geometry*, pages 82–87, New York, NY, USA, 2002. ACM.
- [36] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice Hall, 3rd edition edition, 2008.
- [37] G. Healey and D. Slater. Global color constancy: recognition of objects by use of illumination-invariant properties of color distributions. *Journal of the Optical Society of America*, 11(11):3003–3010, 1994.
- [38] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(3):334–352, Aug. 2004.
- [39] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld. Detection and Location of People in Video Images Using Adaptive Fusion of Color and Edge Information. In *Proc. of the IAPR International Conference on Pattern Recognition*, pages 4627–4631, 2000.
- [40] A. K. Jain and P. J. Flynn. *Three-Dimensional Object Recognition Systems*. Elsevier Science Inc, 1993.
- [41] M. Kampel. 3d data retrieval for pottery documentation. In *Multimedia Content Representation, Classification and Security*, volume 4105 of *Lecture Notes in Computer Science (LNCS)*, pages 362–369. Springer Verlag, 2006.
- [42] M. Kampel. Computer aided analysis of ancient coins. In *IAPR-EVA conference on Digital Cultural Heritage - Essential for Tourism*, pages 127–134, 2008.
- [43] M. Kampel, J. Aguilera, D. Thirde, M. Borg, G. Fernandez, H. Wildenauer, and J. Ferryman. 3d object localisation and evaluation from video streams. In F. Lenzen, O. Scherzer, and M. Vincze, editors, *Proc. of the 30th Workshop of the Austrian Association for Pattern Recognition (OAGM/AAPR)*, pages 113 – 122, 2006.

- [44] M. Kampel and R. Huber-Mörk. An intelligent system for image based retrieval and identification of ancient coins. *IEEE Intelligent Systems*, 2008. under review.
- [45] M. Kampel and C. Liska. Zwei computerunterstützte Methoden zur bildhaften Erfassung archäologischer Fundstück. *Forum Archaeologiae*, III(6), 1998. <http://farch.net>.
- [46] M. Kampel, C. Liska, and S. Tosovic. Adaptive 3D Reconstruction of Archaeological Pottery. In *Proc. of IS&T/SPIE Symposium on Machine Vision Applications in Industrial Inspection IX*, volume 4301, pages 42–51, San Jose, CA, 2001. SPIE.
- [47] M. Kampel, H. Mara, and R. Sablatnig. Investigation on traditional and modern ceramic documentation. In G. Sicuranza G. Vernazza, editor, *International IEEE Conference on Image Processing*, volume 2, pages 570 – 573. IEEE Computer Society, 2005.
- [48] M. Kampel and F. Melero. Virtual fragment reconstruction from fragment’s profile. In F. Niccolucci D. Arnold, A. Chalmers, editor, *Proc. of the 4th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage (VAST04)*, pages 75–84, 2003.
- [49] M. Kampel and R. Sablatnig. A Model-based Approach to Range Image Registration. In Markus Vincze, editor, *Robust Vision for Industrial Applications 1999, Proc. of the 23rd Workshop of the Austrian Association for Pattern Recognition (ÖAGM)*, volume 128 of *OCG Schriftenreihe*, pages 109–117. Oldenbourg Wien, München, 1999.
- [50] M. Kampel and R. Sablatnig. On 3D Modelling of Archaeological Sherds. In N. Sarris and G. Strinzis, editors, *Proc. of International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging*, pages 95–98, Santorini, Greece, 1999.
- [51] M. Kampel and R. Sablatnig. Color Classification of Archaeological Fragments. In A. Sanfeliu, J.J. Villanueva, M. Vanrell, R. Alquezar, A.K. Jain, and J. Kittler, editors, *Proc. of 15th International Conference on Pattern Recognition, Barcelona*, volume 4, pages 771–774. IEEE Computer Society, 2000.
- [52] M. Kampel and R. Sablatnig. Color Classification of Ceramics using Spectral Reflectance. In *First International Conference on Color in Graphics and Image Processing*, pages 323–327, Saint-Etienne, France, October 2000.
- [53] M. Kampel and R. Sablatnig. Automated 3D Recording of Archaeological Pottery. In D. Bearman and F. Garzott, editors, *Proc. of the International*

Conference on Cultural Heritage and Technologies in the Third Millennium (ICHIM01), volume 1, pages 169–182, 2001.

- [54] M. Kampel and R. Sablatnig. On using 3d multimedia tools to measure, reconstruct and visualize an archaeological site. *Forum Archaeologiae, Zeitschrift für klassische Archäologie*, 19(IV), 2001. <http://farch.net>.
- [55] M. Kampel and R. Sablatnig. Automated segmentation of archaeological profiles for classification. In R. Kasturi, D. Laurendeau, and C. Suen, editors, *Proc. of 16th International Conference on Pattern Recognition, Quebec City*, volume 1, pages 57–60. IEEE Computer Society, 2002.
- [56] M. Kampel and R. Sablatnig. Computer aided classification of ceramics. In F. Niccolucci, editor, *Proc. of the 2nd International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage (VAST02), Arezzo*, pages 77–82, Oxford, 2002. Archaeopress.
- [57] M. Kampel and R. Sablatnig. An automated pottery archival and reconstruction system. *Journal of Visualization and Computer Animation*, 14(3):111–120, July 2003.
- [58] M. Kampel and R. Sablatnig. Profile based pottery reconstruction. In D. Martin, editor, *Proc. of IEEE/CVPR Workshop on Applications of Computer Vision in Archaeology*, pages CD-ROM. IEEE Computer Society, 2003.
- [59] M. Kampel and R. Sablatnig. 3d puzzling of archaeological fragments. In *Proc. of 9th Computer Vision Winter Workshop*, pages 31 – 40, 2004.
- [60] M. Kampel and R. Sablatnig. Automatisierte Dokumentation von Keramikfunden aus Palpa auf Grundlage eines 3d-Erfassungssystems. In *Broschüre zur Feldkonferenz des Projektverbundes NASCA: Entwicklung und Adaption Archäometrischer Techniken zur Erforschung der Kulturgeschichte*, pages 39 – 42, 2004.
- [61] M. Kampel and R. Sablatnig. On 3D mosaicing of rotationally symmetric ceramic fragments. In J. Kittler, M. Petrou, and M. Nixon, editors, *Proc. of 17th International Conference on Pattern Recognition, Cambridge, UK*, volume 2, pages 265–268. IEEE Computer Society, 2004.
- [62] M. Kampel and R. Sablatnig. Rule based system for archaeological pottery classification. *Pattern Recognition Letters*, 28(6):740–747, 2007.
- [63] M. Kampel, R. Sablatnig, and E. Costa. Classification of archaeological fragments using profile primitives. In Stefan Scherer, editor, *Computer Vision, Computer Graphics and Photogrammetry - A Common Viewpoint, Proc. of the 25th Workshop of the Austrian Association for Pattern Recognition (OEAGM)*, volume 147 of *Schriftenreihe der OCG*, pages 151–158, Oldenburg, Wien, München, 2001.

- [64] M. Kampel, R. Sablatnig, and H. Mara. Automated documentation system of pottery. In N. Magnenat-Thalmann and J.H. Rindel, editors, *Proc. of 1st International Workshop On 3D Virtual Heritage, Geneva, Switzerland*, pages 14–20, 2002.
- [65] M. Kampel, R. Sablatnig, and H. Mara. Robust 3d reconstruction of archaeological pottery based on concentric circular rills. In *Proc. of the 6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS05)*, pages 14 – 20, 2005.
- [66] M. Kampel, R. Sablatnig, and S. Tosovic. Volume based reconstruction of archaeological artifacts. In W. Boehler, editor, *Proc. of International Workshop on Scanning for Cultural Heritage Recording*, pages 76–83, 2002.
- [67] M. Kampel, S. Tosovic, and R. Sablatnig. Octree-based fusion of shape from silhouette and shape from structured light. In G.M. Cortelazzo and C. Guerra, editors, *Proc. of 1st IEEE International Symposium on 3D Data Processing Visualization and Transmission, Padova*, pages 754–757, 2002.
- [68] M. Kampel, H. Wildenauer, P. Blauensteiner, and A. Hanbury. Improved motion segmentation based on shadow detection. *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, 6(3):1–12, 2007.
- [69] M. Kampel and M. Zaharieva. Optical recognition of medieval coins. In *35th Conference on Computer Applications and Quantitative Methods in Archaeology (CAA '07)*, 2007. in press.
- [70] M. Kampel and M. Zaharieva. Recognizing ancient coins based on local features. In G. Bebis et al., editor, *Advances in Visual Computing*, volume 5358 of *Lecture Notes in Computer Science (LNCS)*, pages 11–22. Springer Berlin / Heidelberg, December 2008. in press.
- [71] A. Karasik and U. Smilansky. 3d scanning technology as a standard archaeological tool for pottery analysis: practice and theory. *Journal of Archaeological Science*, 35(5), 2008.
- [72] R. Khan, J. Stöttinger, and M. Kampel. An adaptive multiple model approach for fast content-based skin detection in on-line videos. In *Proc. of the 1st ACM International Workshop in Analysis and Retrieval of Events/Actions and Workflows in Video Streams*, Vancouver, CN, Oct. 31st 2008. to appear.
- [73] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi. Recent advances in visual and infrared face recognition: a review. *Computer Vision Image Understanding*, 97(1):103–135, 2005.

- [74] L. J. Latecki, R. Lakämper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR00)*, volume 1, pages 424–429, Los Alamitos, CA, USA, 2000. IEEE Computer Society.
- [75] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1265–1799, 2005.
- [76] B. Leibe, N. Cornelis, K. Cornelis, and L. Van Gool. Dynamic 3d scene analysis from a moving vehicle. *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR07*, pages 1–8, June 2007.
- [77] H. C. G. Leitão and J. Stolfi. Measuring the Information Content of Fracture Lines. *International Journal of Computer Vision*, 65(3):163–174, 2005.
- [78] H. Ling and K. Okada. An efficient earth mover’s distance algorithm for robust histogram comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):840–853, 2007.
- [79] G. Loy and J.O. Eklundh. Detecting symmetry and symmetric constellations of features. In *9th European Conference on Computer Vision*, number 3952 in Lecture Notes in Computer Science (LNCS), pages 508–521. Springer, 2006.
- [80] M. Maggini, S. Melacci, and L. Sarti. Representation of facial features by catmull-rom splines. In Walter G. Kropatsch, Martin Kampel, and Allan Hanbury, editors, *12th International Conference on Computer Analysis of Images and Patterns (CAIP’07)*, volume 4673 of *Lecture Notes in Computer Science (LNCS)*, pages 408–415. Springer, 2007.
- [81] C. Maiza and V. Gaildrat. Automatic classification of archaeological potsherd. In *The Eight International Conference on Computer Graphics and Artificial Intelligence, 3IA2005*, pages 11–12, 2005.
- [82] Y. Manabe, M. Tsujita, and K. Chihara. Measurement of shape and refractive index of transparent object. In *Proc. of the 17th International Conference on Pattern Recognition*, volume 2, pages 843–846, 2004.
- [83] H. Mara, M. Kampel, F. Niccolucci, and R. Sablatnig. Ancient coins and ceramics - 3d and 2d documentation for preservation and retrieval of lost heritage. In F. Remondino and S. El-Hakim, editors, *Proc. of the 2nd ISPRS Intl. Workshop on 3D Virtual Reconstruction and Visualization of Complex Architectures (3D-ARCH 2007)*, volume XXXVI-5/W47 of *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 10–18, 2007.

- [84] H. Mara, M. Kampel, and R. Sablatnig. Preprocessing of 3D-data for Classification of Archaeological Fragments in an Automated System. In F. Leberl and F. Fraundorfer, editors, *Vision with Non-Traditional Sensors, Proc. of the 26th Workshop of the Austrian Association for Pattern Recognition (ÖAGM)*, volume 160 of *Schriftenreihe der OCG*, pages 257–264, 2002.
- [85] I. Marie and H. Qasrawi. Virtual assembly of pottery fragments using moiré surface profile measurements. *Journal of Archaeological Science*, 32(10):1527–1533, 2005.
- [86] D. Marr. *Vision - A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, San Francisco, 1982.
- [87] P. Martinez. Digital Realities and Archaeology: A difficult relationship or a fruitful marriage. In D. Arnold, editor, *Proc. of the International Symposium on Virtual Reality, Archaeology and Cultural Heritage, Athens*, pages 9–15, 2001.
- [88] G. McNeill and S. Vijayakumar. 2D shape classification and retrieval. In *Proc. of International Joint Conference on Artificial Intelligence*, pages 1483–1488, 2005.
- [89] K. Mikolajczyk, B. Leibe, and B. Schiele. Local features for object class recognition. In *IEEE International Conference on Computer Vision*, volume 2, pages 1792–1799, 2005.
- [90] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.
- [91] A. C. Murillo, J. J. Guerrero, and C. Sagüés. Surf features for efficient robot localization with omnidirectional images. In *IEEE International Conference on Robotics and Automation*, pages 3901–3907, 2007.
- [92] V. Nautiyal, S. Nautiyal, M. Naithani, J. Sanjiv, J. Clark, J. Landrum, A. Bergstrom, R. Frovarp, J. Haywley, and D. Eichele. 3d modelling and database of indian archaeological pottery assemblages. In *Proc. of the 31st Conference Computer on Applications and Quantitative Methods in Archaeology*, pages 103–107, 2003.
- [93] R. Pflugfelder. *Self-calibrating Cameras in Video Surveillance*. PhD thesis, Graz University of Technology, 2007.
- [94] J. Ponce, M. Hebert, C. Schmid, and A. Zisserman. *Toward Category-Level Object Recognition*, volume 4170 of *Lecture Notes in Computer Science (LNCS)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

- [95] F. Remondino and S. El-Hakim. Image-based 3d modelling: A review. *The Photogrammetric Record*, 21(115):269–291, 2006.
- [96] R. Sablatnig and M. Kampel. On registering front- and backviews of rotationally symmetric objects. In Franc Solina and Aleš Leonardis, editors, *Proc. of 8th Intl. Conf. on Computer Analysis of Images and Patterns, CAIP99*, volume 1689 of *Lecture Notes in Computer Science (LNCS)*, pages 339–346. Springer, 1999.
- [97] R. Sablatnig and M. Kampel. The Virtual Reconstruction of an Archaeological Site - An Overview of the MURALE Project. In B. Likar, editor, *Proc. of 6th Computer Vision Winter Workshop 2000*, pages 60–70, February 2001.
- [98] R. Sablatnig and M. Kampel. Model-based registration of front- and backviews. *Computer Vision and Image Understanding*, 87(1):90–103, Sep. 2002.
- [99] R. Sablatnig, S. Tosovic, and M. Kampel. Volume computation for archaeological vessels. In B. Mafart and H. Delingette, editors, *Proc. of 14th Congress of the International Union of Prehistoric and Protohistoric Sciences — Three Dimensional Imaging in Paleoanthropology and Prehistoric Archaeology*, volume 1049 of *BAR International Series*, pages 93–99. Archeopress, 2002.
- [100] J. Salvi, C. Matabosch, D. Fofi, and J. Forest. A review of recent range image registration methods with accuracy evaluation. *Image Vision Computing*, 25(5):578–596, 2007.
- [101] S. Savarese, M. Andreetto, H. Rushmeier, F. Bernardini, and P. Perona. 3d reconstruction by shadow carving: Theory and practical evaluation. *International Journal Computer Vision*, 71(3):305–336, 2007.
- [102] N. Setiawan, H. Seok-Ju, K. Jang-Woon, and L Chil-Woo. Gaussian mixture model in improved hls color space for human silhouette extraction. In *Advances in Artificial Reality and Tele-Existence*, volume 4282, pages 732–741. Springer Berlin / Heidelberg, 2006.
- [103] M. Shah and R. Jain. *Motion-Based Recognition*. Kluwer Academic Publisher, 1997.
- [104] L.G. Shapiro and G Stockman. *Computer Vision*. Prentice Hall, 2001.
- [105] J. Shi and C. Tomasi. Good features to track. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.
- [106] M. Smith and L. Smith. Computer vision applications - special issue. *Image and Vision Computing*, 25(7):1035–1036, 2007.

- [107] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. CENGAGE-Engineering, third edition, 2007.
- [108] C. Stauffer and W.E.L. Grimson. Adaptive Background Mixture Models for Real-Time Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 246–252, 1999.
- [109] G. D. Sullivan. Visual interpretation of known objects in constrained scenes. In *Philosophical Transactions, Royal Society London*, volume B337, pages 361–370, 1992.
- [110] L. Szumilas. *Scale and Rotation Invariant Shape Matching*. PhD thesis, Vienna University of Technology, 2008.
- [111] D. Thirde, M. Borg, J. Aguilera, J. Ferryman, K. Baker, and M. Kampel. Evaluation of object tracking for aircraft activity surveillance. In T. Tan R. Chellappa, J. Ferryman, editor, *In Proc. of the Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Beijing*, pages 145 – 152, 2005.
- [112] D. Thirde, M. Borg, J. Aguilera, H. Wildenauer, J. Ferryman, and M. Kampel. Robust real-time tracking for visual surveillance. *EURASIP Journal on Advances in Signal Processing*, 2007:23, 2007.
- [113] D. Thirde, M. Borg, J. Ferryman, J. Aguilera, and M. Kampel. Distributed multi-camera surveillance for aircraft servicing operations. In *Advances in Visual Computing*, volume 3804/2005 of *Lecture Notes in Computer Science (LNCS)*, pages 118–125, 2005.
- [114] D. Thirde, M. Borg, V. Valentin, L. Bartelemy, J. Aguilera, G. Fernandez, J. Ferryman, F. Bremond, and M. Kampel. People and vehicle tracking for visual surveillance. In *6th IEEE International Workshop on Visual Surveillance*, pages 169–176, 2006.
- [115] D. Thirde, M. Borg, V. Valentin, F. Fusier, J. Aguilera, J. Ferryman, F. Bremond, M. Thonnat, and M. Kampel. Visual surveillance for aircraft activity monitoring. In T. Tan R. Chellappa, J. Ferryman, editor, *In Proc. of the Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 255 – 262, 2005.
- [116] S. Todorovic and N. Ahuja. Extracting subimages of an unknown category from a set of images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR06)*, pages 927–934. IEEE Computer Society, 2006.
- [117] T. Tuytelaars and L. V. Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 59(1):61–85, 2004.

- [118] H. Wildenauer, J. Aguilera, and M. Kampel. An evaluation of error metrics for the segmentation of moving objects on an airport’s apron. In M. Vincze D. Chetverikov, L. Czuni, editor, *In Proc. of the Joint Hungarian-Austrian Conference on Image Processing and Pattern Recognition*, pages 119 – 126, 2005.
- [119] H. Wildenauer, P. Blauensteiner, A. Hanbury, and M. Kampel. Motion detection using an improved colour model. In *Advances in Visual Computing*, volume 4292 of *Lecture Notes in Computer Science (LNCS)*, pages 607–616. Springer Berlin / Heidelberg, 2006.
- [120] A. Willis. *Stochastic 3D Geometric Models for Classification, Deformation, and Estimation*. PhD thesis, Brown University, 2004.
- [121] A. Willis, X. Orriols, and D. B. Cooper. Accurately estimating sherd 3d surface geometry with application to pot reconstruction. *Computer Vision and Pattern Recognition Workshop*, 1:5, 2003.
- [122] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.
- [123] M. Zaharieva, R. Huber-Mörk, M. Nölle, and M. Kampel. On ancient coin classification. In David Arnold, Alan Chalmers, and Franco Niccolucci, editors, *8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST’07)*, pages 55–62. Eurographics, November 2007.
- [124] M. Zaharieva, M. Kampel, and K. Vondrovec. From manual to automated optical recognition of ancient coins. In *13th International Conference on Virtual Systems and Multimedia (VSMM’07)*, volume 4820/2008 of *Lecture Notes in Computer Science (LNCS)*, pages 88–99. Springer, September 2007.
- [125] M. Zaharieva, M. Kampel, and S. Zambanini. Image based recognition of ancient coins. In *12th International Conference on Computer Analysis of Images and Patterns (CAIP’07)*, volume 4673/2007 of *Lecture Notes in Computer Science (LNCS)*, pages 547–554. Springer, 2007.
- [126] M. Zaharieva, M. Kampel, and S. Zambanini. Image based recognition of coins – an overview of the coins project. In *Proc. of the 31st Workshop of the Austrian Association for Pattern Recognition (AAPR/OAGM)*, pages 57–64, 2007.
- [127] S. Zambanini and M. Kampel. Segmentation of ancient coins based on local entropy and gray value range. In *13th Computer Vision Winter Workshop (CVWW’08)*, April 2008.

- [128] S. Zambanini, M. Kampel, and M. Schlapke. On the use of computer vision for numismatic research. In *International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST08)*, Praga, Portugal, 2008. in press.

Chapter 2

Octree-based Fusion of Shape from Silhouette and Shape from Structured Light

Martin Kampel, Srdan Tosovic, and Robert Sablatnig. Octree-based Fusion of Shape from Silhouette and Shape from Structured Light, In G.M. Cortelazzo and C. Guerra, editors, *Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission (3DPVT02)*, IEEE Computer Society, pp. 754-757, 2002.

Octree-based Fusion of Shape from Silhouette and Shape from Structured Light *

Martin Kampel, Srdan Tosovic, and Robert Sablatnig
Vienna University of Technology,
Institute of Computer Aided Automation,
Pattern Recognition and Image Processing Group
Favoritenstr. 9, 183-2, A-1040 Vienna
{kempel,tos,sab}@prip.tuwien.ac.at, <http://www.prip.tuwien.ac.at>

Abstract

An algorithm for the automatic construction of a 3d model of archaeological vessels using two different 3d algorithms is presented. In archeology the determination of the exact volume of arbitrary vessels is of importance since this provides information about the manufacturer and the usage of the vessel. To acquire the 3d shape of objects with handles is complicated, since occlusions of the object's surface are introduced by the handle and can only be resolved by taking multiple views. Therefore, the 3d reconstruction is based on a sequence of images of the object taken from different viewpoints with different algorithms; shape from silhouette and shape from structured light. The output of both algorithms are then used to construct a single 3d model. Results of the algorithm developed are presented for both synthetic and real input images.

1 Introduction

The combination of the *Shape from Silhouette* (SfS) method with the *Shape from Structured Light* (SfSL) method presented in this paper was performed within the *Computer Aided Classification of Ceramics* [7] project, which aims to provide an objective and automated method for classification and reconstruction of archaeological pottery. Pottery was made in a very wide range of forms and shapes. The purpose of classification is to get a systematic view of the material found, to recognize types, and to add labels for additional information as a measure of quantity [13]. In this context, decoration of pottery is of great interest. Decoration is difficult to illustrate since it is a perspective projection of an originally spherical surface. In order

*This work was partly supported by the Austrian Science Foundation (FWF) under grant P13385-INF, the European Union under grant IST-1999-20273 and the Austrian Federal Ministry of Education, Science and Culture.

to be able to unwrap the surface it is necessary to have a 3d representation of the original surface. Furthermore, the exact volume of the vessel is of great interest to archaeologists too, since the volume estimation allows also a more precise classification [13].

SfS is a method of automatic construction of a 3D model of an object based on a sequence of images of the object taken from multiple views, in which the object's silhouette represents the only interesting feature of the image [16, 15]. The object's silhouette in each input image corresponds to a conic volume in the object real-world space. A 3D model of the object can be built by intersecting the conic volumes from all views, which is also called *Space Carving* [8].

There have been many works on construction of 3D models of objects from multiple views ([1, 10, 4, 18, 15]). Szeliski [16] first creates a low resolution octree model quickly and then refines this model iteratively, by intersecting each new silhouette with the already existing model. Niem [12] uses pillar-like volume elements instead of an octree for the model representation. De Bonet and Viola [3] extended the idea of voxel reconstruction to transparent objects by introducing the Roxel algorithm — a responsibility weighted 3D volume reconstruction. Wong and Cipolla [20] use uncalibrated silhouette images and recover the camera positions and orientations from circular motions.

SfS can be applied on objects of arbitrary shapes, including objects with certain concavities (like a handle of a cup), as long as the concavities are visible from at least one input view. This condition is very hard to hold since most of the archaeological vessels do have concavities. To discover these concavities we use SfSL, which is based on active triangulation [2, 6]. Most laser light based SfSL methods use a camera, a calibrated laser ray or plane and a motion platform — usually a linear slide or a turntable.

The work of Szeliski [16] was used as a basis for the SfS and the work of Liska [9] as a basis for the SfSL approach presented in this paper which is organized as follows. Section 2 describes the equipment used for acquisition. Section

3 presents the combination strategy proposed. Experimental results with both synthetic and real data are given in Section 4. At the end of the paper conclusions are drawn and future work is outlined.

2 Acquisition System

The acquisition system, shown in Figure 1, consists of a turntable (diameter 50 cm), two monochrome CCD-cameras ($f=16\text{ mm}$, 768×576 pixels), a laser and a lamp. Both cameras are placed in a distance of about 50 cm from the rotational axis of the turntable. Ideally the optical axis of the camera for acquiring object's silhouettes (Camera-1 in Figure 1) lies nearly in the rotational plane of the turntable, orthogonal to the rotational axis. The camera for acquiring the projection of the laser plane onto the object (Camera-2 in Figure 1) views the turntable in an angle of about 45° . The laser is directed such that the light plane it projects contains the rotational axis of the turntable. Camera-2 from Figure 1 views the light plane also from an angle of about 45° . The relative position of the two cameras to one another is not important, since the acquisition of the silhouettes and the acquisition of the laser light projection are independent from one another.

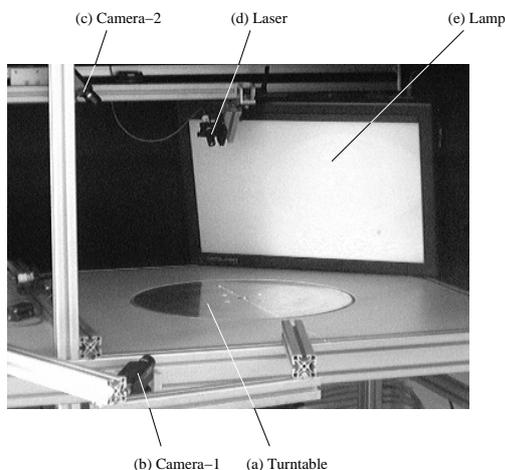


Figure 1. Acquisition system

Prior to any acquisition, the system is calibrated in order to determine the inner and outer orientation of the camera and the rotational axis of the turntable. We used the calibration technique proposed by Roger Y. Tsai [17], for several reasons: it is efficient and accurate, lens distortion can be taken into account but also ignored if desired, and there is a publicly available implementation [19]. In our experiments, the average calibration error was 0.5 pixel or less (measured

in the image plane), which is sufficient for our approach, because the smallest unit processed in an image is 1 pixel.

3 Fusion of Algorithms

The first step between the image acquisition and creation of the final 3D model of an object consists of converting the images acquired into binary images. A pixel in such a binary image should have the value 0 if it represents a point in 3D space which does not belong to the object *for sure*, and the value of 1 otherwise. The binarization is performed on input images for both SfS and SfSL.

For the SfS part of the method presented, a reliable extraction of the object's silhouette from an acquired image is of crucial importance for obtaining an accurate 3D model of an object. If the background brightness is not uniform, the silhouette extraction can be a difficult task. For that reason, in addition to the images of the object (Figure 2a, upper image) taken from different viewpoints, an image of the acquisition space is taken, without any object in it. Then, the absolute difference between this image and an input image is built, which creates an image with a uniform background and a high contrast between the object and the background. Next, thresholding is used to create a binary image (Figure 2a, upper image) where pixels with the value 1 represent the object's silhouette and those with value 0 the background.

Another option for extracting object silhouettes from input images would be to use edge detection [11] instead of thresholding. This approach could be more accurate, even a sub-pixel precision could be reached, but it is also more complex.

An input image for SfSL contains the projection of a laser plane onto the object (Figure 2a, lower image). A white pixel in this image represents a 3D point on the object's surface which intersects the laser plane. A black pixel represents a 3D point in the laser plane which does not belong to the object's surface — it is either inside the object or it does not belong to the object at all. The creation of a SfSL binary image is more complex. Based on the known position of the laser, an input image (Figure 2a, lower left image) is converted to an image approximating intersection of the laser plane with the whole object (Figure 2a, lower right image).

Our approach builds a 3D model of an object performing the following steps (illustrated in Figure 2): First, both of the input images (SfS and SfSL) are binarized such that the white image pixels *possibly* belong to the object and the black pixels *for sure* belong to the background (Figure 2a). Then, the initial octree containing one single root node marked "black" is build (Figure 2b). Black nodes are subsequently checked by projecting the nodes into all SfS binarized input images and intersecting them with the im-

age silhouettes of the object (Figure 2c). As the result of the intersection the node can remain "black" (if it lies within the object) or be set to "white" (if it lies outside the object) or "grey" (if it lies partly within and partly outside the object). If the resulting node is not white, it is projected into the binarized SfSL image representing the nearest laser plane to the node and again intersected. All grey nodes are divided into 8 child nodes all of which are marked "black" and the intersection test is performed in each of the black nodes. This subdivision of grey nodes is done until there are no grey nodes left or a subdivision is not possible (voxel size), which results in the final model (Figure 2d).

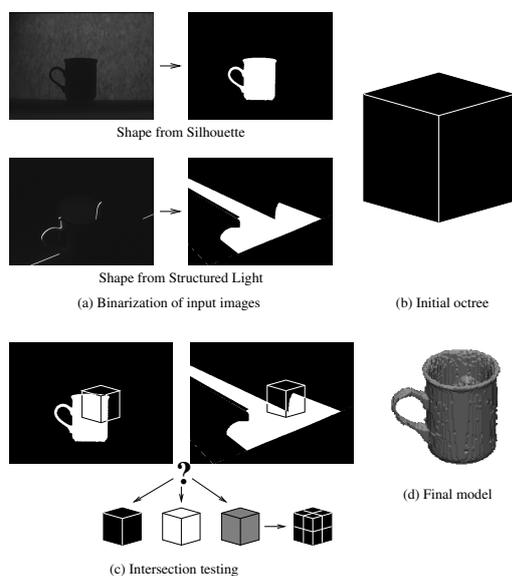


Figure 2. Algorithm overview

4 Results

Experiments were performed with both synthetic and real objects. For synthetic objects we built a model of a virtual camera and laser and created input images such that the images fit perfectly into the camera model. For both synthetic and real objects we compare the volume and the size of the bounding cuboid of the model with the volume and size of the bounding cuboid of the object. As synthetic objects we created a virtual sphere with the radius 200 mm, and a virtual cuboid with dimensions 100 × 70 × 60 mm. The images of the sphere were constructed such that both SfS and SfSL alone can reconstruct the object completely, whereas for the cuboid a more realistic case was simulated, where the structured light images contain occlusions. The models of these objects were constructed with different parameter values, such as the number of views used and the

maximal octree resolution.

The tests with the sphere showed that SfS and SfSL perform similarly when they have perfect input images — starting from resolution 128³, both methods were able to create models with the approximation error of 2% or less. Regarding the number of views, 20 views were sufficient for both methods in order to create models with the volume less than 1% different from the models built using 360 views. With the synthetic cuboid, neither of the methods was able to reconstruct the cuboid completely, but the combined method constructed its perfect model starting from the resolution 128³. However, even if using 180 views instead of 360, the volume error of the cuboid was greater than 1% (1.45%), which indicates that flat surfaces are more difficult to model with our method. Table 1 summarizes the results of the models built.

For tests with real objects we used 8 objects: a metal cuboid, a wooden cone, a globe, a coffee cup, two archaeological vessels and two archaeological sherds. The real volume of the first 3 objects can be computed analytically. For the two vessel it could be theoretically measured by putting water into the objects, but it has not been done since the vessels do have holes, which we are not allowed to close, so for these objects we can only compare the bounding cuboid of the model and the object. Figure 3 shows the objects and their models built using 360 views for each of the underlying methods and the octree resolution 256³.

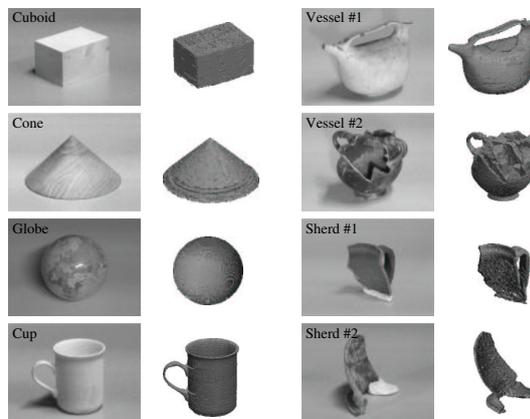


Figure 3. Real objects and their models

The error of the computed volume for real objects was between 3% and 13%, by an order of magnitude larger than the errors with synthetic objects. The main reason turned out to be the threshold based binarization of silhouette images, which interpreted parts of the object as the background, especially close to the turntable surface. That explains why the error was the largest for the cone and the smallest for the globe (see Table 1). The cone has a large base leaning on the turntable, while the globe only touches

the turntable in an almost tangential way.

object	octree	#views	volume	vol.error
synth. sphere	—	analytic	33 510 322	—
	64 ³	360+360	35 241 984	+5.17%
	128 ³	360+360	33 786 880	+0.83%
	256 ³	360+360	33 034 528	-1.42%
	256 ³	180+180	33 067 552	-1.32%
synth. cuboid	—	analytic	420 000	—
	64 ³	360+360	432 000	+2.86%
	128 ³	360+360	420 000	0.00%
	256 ³	360+360	420 000	0.00%
	256 ³	180+180	426 071	+1.45%
real cuboid	—	analytic	420 000	—
	256 ³	360+360	384 678	-8.41%
cone	—	analytic	496 950	—
	256 ³	360+360	435 180	-12.43%
globe	—	analytic	1 756 564	—
	256 ³	360+360	1 717 624	-2.22%
cup	—	analytic	N/A	—
	256 ³	360+360	276 440	N/A
vessel #1	—	analytic	N/A	—
	256 ³	360+360	336 131	N/A
vessel #2	—	analytic	N/A	—
	256 ³	360+360	263 696	N/A
sherd #1	—	analytic	N/A	—
	256 ³	360+360	35 911	N/A
sherd #2	—	analytic	N/A	—
	256 ³	360+360	38 586	N/A

Table 1. Volume of objects and their models

5 Conclusion

This paper presented a 3D modeling method based on combination of SfS and laser based SfSL, using a turntable to obtain multiple views of an object. The purpose of combining SfS and SfSL was to create a method which will use the advantages and overcome the weaknesses of both underlying methods and create complete models of arbitrarily shaped objects. The experiments with synthetic objects showed that construction of nearly perfect models is possible, limited only by image and model resolution. In the experiments with real objects the results were less accurate, but the algorithm was able to produce complete and visually faithful models for all objects, including sherds and vessels with concave surfaces and a handle.

Overall, our combined modeling approach proved to be useful for automatic creation of models of arbitrarily shaped objects. With respect to its archaeological application it can provide models of any kind of archaeological pottery. Furthermore, the volume of an object can be estimated, including the inside volume of objects such as bowls or cups. However, for high precision measurements of the volume our method did not produce highly accurate results, but it gave a good rough estimate, which is sufficient for most archaeological applications. Higher accuracy could be achieved by improving the binarization of input images, which showed to be the main reason for relatively large errors for real objects. A possible enhancement to our method

would be to take additional color images of an object and perform texture mapping onto the model, which would improve the visual impression of the models built.

References

- [1] H. Baker. Three-dimensional modelling. In *Proc. of IJCAI*, pages 649–655, 1977.
- [2] P. Besl. Active, optical range imaging sensors. *MVA*, 1(2):127–152, 1988.
- [3] J. S. D. Bonet and P. Viola. Voxels: Responsibility weighted 3D volume reconstruction. In *Proc. of 7th ICCV*, pages 418–425, 1999.
- [4] C. H. Chien and J. K. Aggarwal. Volume/surface octrees for the representation of three-dimensional objects. *CVGIP*, 36:100–113, 1983.
- [5] J. Davis and X. Chen. A laser range scanner designed for minimum calibration complexity. In *Proc. of 3rd Intl. Conf. on 3-D Digital Imaging and Modeling*, pages 91–98, 2001.
- [6] F. DePiero and M. Trivedi. 3-D computer vision using structured light: Design, calibration, and implementation issues. *Advances in Computers*, 43:243–278, 1996.
- [7] M. Kampel and R. Sablatnig. Automated 3D recording of archaeological pottery. In *D. Bearman and F. Garzotto (Ed.). Proc. of the Intl. Conf on Cultural Heritage and Technologies in the Third Millennium Milan, Italy*, pages 169–182, 2001.
- [8] K. Kutulakos and S. Seitz. A theory of shape by space carving. *IJCV*, 38(3):197–216, July 2000.
- [9] C. Liska and R. Sablatnig. Estimating the next sensor position based on surface characteristics. In *Proc. of ICPR*, volume I, pages 538–541, 2000.
- [10] W. N. Martin and J. K. Aggarwal. Volumetric description of objects from multiple views. *IEEE-PAMI*, 5(2):150–158, 1983.
- [11] V. S. Nalwa and T. O. Binford. On detecting edges. *IEEE-PAMI*, 8:699–714, November 1986.
- [12] W. Niem. Robust and fast modelling of 3D natural objects from multiple views. In *Image and Video Processing II, Proc. of SPIE*, pages 388–397, 1994.
- [13] C. Orton, P. Tyers, and A. Vince. *Pottery in Archaeology*, 1993.
- [14] J. Park, G. N. DeSouza, and A. C. Kak. Dual-beam structured-light scanning for 3-D object modeling. In *Proc. of 3rd Intl. Conf. on 3-D Digital Imaging and Modeling*, pages 65–72, 2001.
- [15] M. Potmesil. Generating octree models of 3D objects from their silhouettes in a sequence of images. *CVGIP*, 40:1–29, 1987.
- [16] R. Szeliski. Rapid octree construction from image sequences. *CVGIP-IU*, 58(1):23–32, July 1993.
- [17] R. Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proc. of CVPR*, pages 364–374, 1986.
- [18] J. Veenstra and N. Ahuja. Efficient octree generation from silhouettes. In *Proc. of CVPR*, pages 537–542, 1986.
- [19] R. G. Willson. Tsai camera calibration software. <http://www.cs.cmu.edu/~rgw/TsaiCode.html>.
- [20] K. Y. K. Wong and R. Cipolla. Structure and motion from silhouettes. In *Proc. of ICCV*, pages 217–222, 2001.

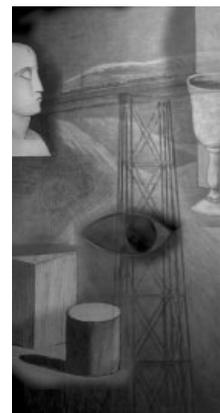
Chapter 3

An Automated Pottery Archival and Reconstruction System

Martin Kampel and Robert Sablatnig. An Automated Pottery Archival and Reconstruction System, *Journal of Visualization and Computer Animation*, John Wiley & Sons, Ltd., Vol. 14, pp. 111-120, 2003.

An automated pottery archival and reconstruction system

By Martin Kampel* and Robert Sablatnig



Motivated by the current requirements of archaeologists, we are developing an automated archival system for archaeological classification and reconstruction of ceramics. Our system uses the profile of an archaeological fragment, which is the cross-section of the fragment in the direction of the rotational axis of symmetry, to classify and reconstruct it virtually. Ceramic fragments are recorded automatically by a 3D measurement system based on structured (coded) light. The input data for the estimation of the profile is a set of points produced by the acquisition system. By registering the front and the back views of the fragment the profile is computed and measurements like diameter, area percentage of the complete vessel, height and width are derived automatically. We demonstrate the method and give results on synthetic and real data. Copyright © 2003 John Wiley & Sons, Ltd.

Received: 29 November 2002; Revised 19 March 2003

KEY WORDS: pottery, 3D registration; pottery reconstruction; pottery archival; documentation

Introduction

New technologies are introduced to old research areas and provide new insights for both the researchers and people interested in this field. The truth of this statement can be demonstrated especially in the field of archaeology, since there are many researchers in that area who already use new technologies, and many people have become interested in the field of archaeology since so-called Archaeo-Parks became popular with visitors.^{1,2}

Ceramics are among the most widespread archaeological finds, having a short period of use. A large number of ceramic fragments are found at nearly every excavation (Figure 1) and have to be photographed, measured, drawn and classified.

Because conventional documentation methods have been shown to be unsatisfactory,³ the interest in finding any automatic solution has increased.⁴ Cooper *et al.*⁵ present an approach to a largely automated estimation

of polynomial models in order to assemble virtual pots from 3D measurements of their fragments. Existing techniques on the fragment reconstruction problem mainly focus on the analysis of the break curve.⁶ In particular, Copper *et al.*⁷ developed a method for fragment matching based on a Bayesian approach using break curves, estimated axis and profile curves. Kong and Kimia⁸ try to solve the jigsaw problem in two stages: first, pot sherds are joined automatically in two dimensions by using an efficient joint detection algorithm. Next, three-dimensional shape is recovered by adequate three-dimensional transformation. Leitao and Stolfi⁹ describe an algorithm for reassembling broken two-dimensional fragments. The procedure compares the curvature-encoded fragment outlines.

Our approach to pottery reconstruction, in contrast, concentrates on the virtual reconstruction of vessels out of one (large enough) fragment; the reassembling of broken vessels from fragments is not considered in this paper. The reconstruction can only be performed if *a priori* knowledge on the type and class of vessel of which the fragment is a part is provided. This fact forces us to have a profile-based classification strategy that integrates the archaeological knowledge of types and forms of vessels in a specific regional area. The classification scheme itself is provided by archaeologists and is also designed as a tool for them in order to produce correct virtual reconstructions. Mimicking the archaeological approach to pottery classification, our automated approach is based on estimation of the correct

*Correspondence to: Martin Kampel, Pattern Recognition and Image Processing Group, Vienna University of Technology, Favoritenstraße 9/1832, A-1040 Vienna, Austria.
E-mail: kampel@prip.tuwien.ac.at

Contract/grant sponsor: Austrian Science Foundation; contract/grant number: P13385-INF.
Contract/grant sponsor: European Union; contract/grant number: IST-1999-20273.
Contract/grant sponsor: Austrian Federal Ministry of Education, Science and Culture.



Figure 1. Boxes filled with ceramics stored in archives.

orientation of the fragment, which leads to the exact position of a fragment on the original vessel. In the reconstruction phase partial similarities of profiles can be detected and complete pots can be reconstructed based on the previously stored data. Figure 2 shows the automated archival process schematically, giving an overview of the intermediate steps.

The range and pictorial information of a fragment provided by the acquisition system is the basis for the further documentation and classification process.¹⁰ Using the data of the fragment's inner and outer surface, the axis of rotation has to be determined for both surfaces. Once the axes are determined the profile section can be generated by registering the two surfaces on one another. Next, the longest intersection of a plane that is rotated within the rotational axis of the fragment with the surfaces of the fragment is determined. The profile section of the fragment, which is the cross-section of the fragment in the direction of the rotational axis of symmetry, is the result of the documentation step. With the help of the profile, measurements like diameter, area, percentage of complete vessel, height,

width, etc. are computed automatically. Next a classification process tries to find matching profile parts of already classified vessel shapes in order to reconstruct the complete vessel shape.¹¹ Note that the shape object is virtually reconstructed without putting together all fragments.

Following the manual strategy of archaeologists, the profile is first segmented automatically into its parts, the so-called *primitives*. Our approach is a hierarchical segmentation of the profile into rim, wall and base by segmentation rules based on expert knowledge of the archaeologists and the curvature of the profile.^{12,13} The segments of the curve are divided by so-called segmentation points. Our formalized approach uses mathematical curves to find the extremal and inflection points necessary to classify the original fragment. It is based on cubic B-splines.¹⁴

This paper focuses on the documentation part, in particular on the automatic orientation of fragments, followed by extraction of the profile line. Late-Roman burnished ware, which was found during the excavations from 1968 to 1977 in the legionary fortress of Carnuntum,¹⁵ was chosen as test data for our research.¹⁶

The paper is organized as follows: in the next section we first describe the algorithm for finding the correct orientation automatically. Next the generation of the profile is shown in the third section. In the fourth section results are presented and in the final section we conclude with an outlook on future research.

Fragment Orientation

The acquisition method for estimating the 3D shape of a sherd is shape from structured light,¹⁷ which is based on active triangulation. The projector projects stripe patterns onto the surface of the objects. In order to distinguish between stripes they are binary coded. The

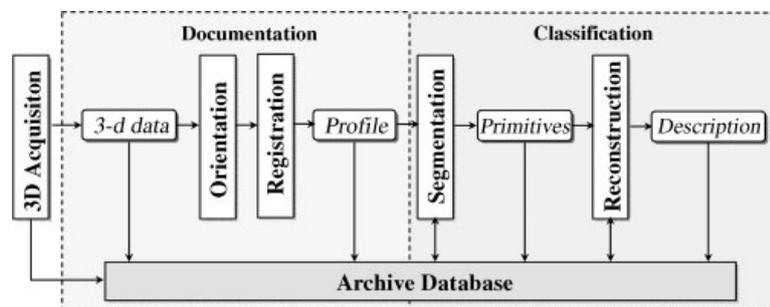


Figure 2. Automated archival process.

camera grabs grey-level images of the distorted light patterns at different times. With the help of the code and the known orientation parameters of the acquisition system, the 3D information of the observed scene point can be computed.¹⁸ The image obtained is a 2D array of depth values called a range image. For on-site recording a portable 3D sensor (Eyetronec ShapeCam¹⁹) was used.

Archaeological pottery is assumed to be rotationally symmetric since it was made on a rotation plate. With respect to this property the axis of rotation is calculated using a Hough-inspired method.²⁰ To perform the registration of the two surfaces of one fragment, we use *a priori* information about fragments belonging to a complete vessel: both surfaces have the same axis of rotation since they belong to the same object. Potmann *et al.*²¹ proposed a solution to reconstruct helical surfaces or surfaces of revolution using line geometric concepts. Their algorithm is based on the fact that the normals of the surfaces lie in linear complexes. Our estimation of the axes of rotation exploits the fact that surface normals of rotational symmetric objects intersect their axis of rotation. The basis for this axis estimation consists of dense range images provided by the range sensor. If we have an object of revolution, like an archaeological vessel made on a rotation plate, we can suppose that all intersections n_i of the surface normals are positioned along the axis of symmetry a .

For each point on the object the surface normal has to be computed. A planar patch of size $s \times s$ can be fitted to the original data using the Minor Component Analysis,²² which minimizes the distance between the points of the surface and the planar patch in an iterative manner in order to compute the optimal value of the normal and discard outliers. For each point on the object, the surface normals n_i are computed using Minor Component Analysis. In order to determine the axis of rotation a all surface normals n_i are clustered in a 3D Hough space: all the points belonging to a line n_i are incremented in the accumulator. Hence the points belonging to a large number of lines (like the points along the axis) will have high counter values. All points in the accumulator with a high counter value are defined as maxima.

In the next step the line formed by the maxima has to be estimated. There are different techniques to solve this problem; in our case the PCA or principal component analysis²³ is used. We have some *a priori* knowledge, namely that the maxima are distributed according to a line (the axis of rotation). The PCA will determine the axis of maximal variance, which is in fact the axis of rotation. The accumulator maxima are taken as candi-

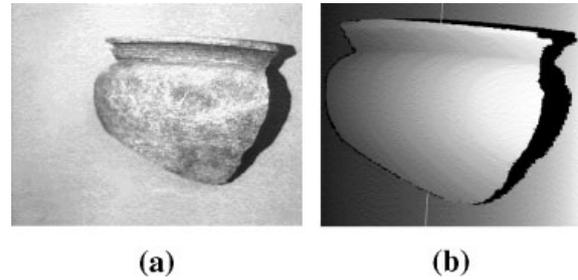


Figure 3. Intensity image (a) and range image (b) of a fragment with rotational axis of the front view.

date points for the estimation of the axis of rotation. Using this technique outliers introduced by noisy range data or discretization errors can be avoided, since in Hough space wrong data points are in the minority and do not build a maximum.

Figure 3(a) shows the intensity image for the front view of a fragment; Figure 3(b) shows the result on the range image with the estimated rotational axis (black regions in Figure 3(b) indicate points where no range information is available due to occlusion of light stripes).

To perform the registration of the two surfaces, we know that both surfaces have the same axis of rotation and the distance of the inner surface to the axis of rotation is smaller than the distance of the outer surface. Finally, both surfaces should have approximately the same profile; i.e. the thickness of the fragment measured on a plane perpendicular to the rotational axis should be constant on average.

The goal of the registration is to find the transformation that relates these two views to one another, thus bringing them into alignment so that the two surfaces represent the object in 3D.²⁴ The most commonly used algorithm for registration is the ICP algorithm.²⁵ ICP iteratively improves the registration of two overlapping surfaces by calculating the unique transformation that minimizes the mean square distances of local point correspondences between the two surfaces (see reference 26 for a review). We register the range images by calculating the axis of rotation of each view (Figures 4a and 4b) and by bringing the resulting axes into alignment (Figure 4c). Knowing the surface normals of all surface patches we transform them into a common reference coordinate system. The first rough alignment is performed by aligning the two surfaces vertically. To do so we select the 10 uppermost points of each surface (we take the uppermost points since rim fragments are the most important fragments in archaeology and they

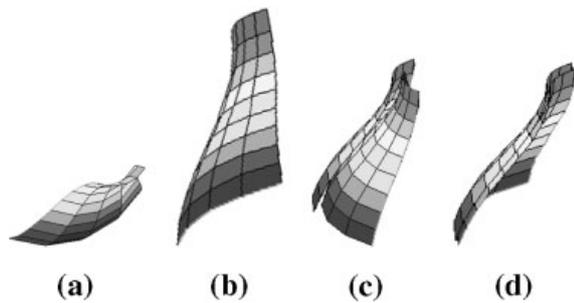


Figure 4. Registration steps using synthetic data.

have the property that all points of the rim lie in the plane perpendicular to the rotational axes) and align them vertically. Next we perform the horizontal alignment by rotating one surface relative to the other until both surfaces have a maximum number of points in a common projection normal to the fixed surface.

In the next step we have to align the surfaces of the objects to avoid intersecting surfaces. The correct match is calculated using a slightly modified ICP algorithm.²⁷ The difference to the standard ICP is that we are calculating the unique transformation that minimizes the mean square distances of the correspondences between the two surfaces to a constant value instead of to zero. This distance d_n is the distance of the two surfaces on a plane perpendicular to the rotational axis, where n denotes the vertical position on the axis. Corresponding points of the two surfaces are estimated by computing the Euclidean distance of the candidate points on the inner surface to the normal on the rotational axis for the point on the outer surface. The point with the minimal distance is taken as corresponding point.

Next the ICP starts by iteratively minimizing the error δ_i , which is the mean error of the local surface distances to d_n until all δ_i are positive (i.e. surfaces do not intersect). Then all d_n are updated to the mean distance of the surfaces in the direction of the rotational axis, the mean square error δ of the local surface distances are computed and the process is restarted. The algorithm ends if there are no significant improvements or the overall error increases. A detailed description of the registration algorithm can be found in reference 28.

Figure 4(e) shows the result for synthetic range data with 50 surface points for each view and a distance of 2.9 mm. The computed distance between the inner and the outer surface is 2.9 mm. The registration error is $\delta = 0.05$ mm; the mean square errors between the original and the computed axes are 0.26 mm and 0.31 mm respectively. Figure 5 shows the registration of inter-

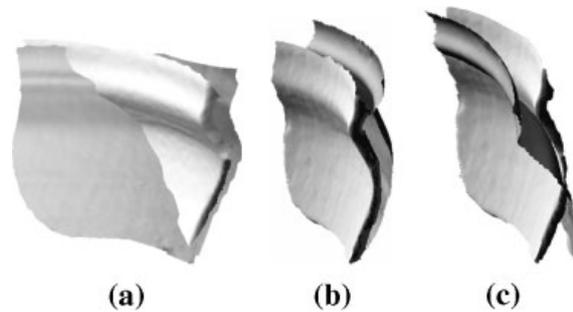


Figure 5. Registration steps using real data.

secting surfaces for real data in detail: Figures 5(a) and 5(b) show intersecting surfaces due to wrong rotational axis estimation; Figure 5(c) shows the same surfaces after the ICP-based registration procedure.

The evaluation of the method (for details see reference 28) shows that the quality of the result is influenced by the number of points in the two views (resolution of the 3D scanner and the object shape for occlusions). The next parameter that influences the results is the mean curvature. Since the registration algorithm uses the axis of rotation for rough alignment, surfaces that are flat cannot be registered since rough alignment does not work. The maximum error was approximately 25%, which is acceptable in this specific application only since the main goal is to compute the profile and the outer profile is the most important attribute. However, the average error of 18% was accepted by the archaeologists since manual drawings and measurement have more errors.³ The rendered 3D models are also used for visualization, as can be seen in Figure 6. Figures 6(a), 6(b) and 6(c) show two more fragments as examples.

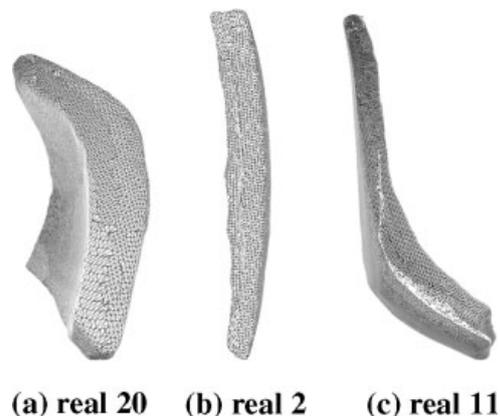


Figure 6. Visual results for registered fragments.

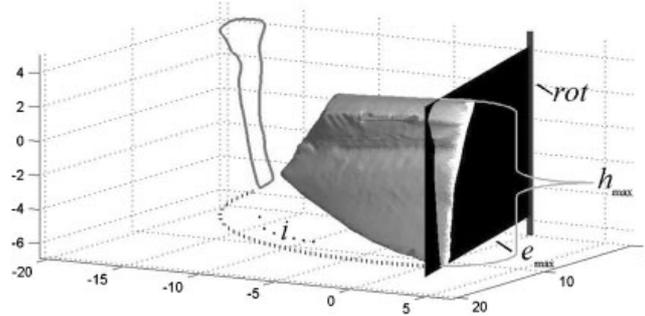


Figure 7. Oriented sherd, rotational axis rot , intersecting plane e_{max} and longest profile line.

Profile Estimation

The registration of front and back view together with the axis of rotation provide the profile used to reconstruct the vessel. Figure 7 shows the 3D model of a sherd and its rotational axis rot as a vertical line along the z -axis. The black plane represents the intersecting plane e_{max} at the maximum height h_{max} of the sherd. The longest profile line is supposed to be the longest elongation along the surface of the sherd parallel to the rotational axis rot . The extracted profile line is shown in the xz -plane. Our algorithm for the estimation of the longest profile line consists of the following steps:

1. First the axis of rotation is transformed into the z -axis of the coordinate system in order to simplify further computation.
2. The fragment's size is calculated by its arc length. Depending on the size we compute a number of intersecting planes e_i , which are used for the profile estimation. The number of planes e_i depends on the length of the perimeter of the fragment. Experiments have shown that 7 to 13 profile lines return the best ratio of exactness and performance. Figure 8 shows a sample of four planes e_i intersecting the 3D model and the plots of the extracted profile lines on the surface of the sherd.
3. A profile line is calculated by intersecting the 3D data of the fragment with planes e_i : first the distance of each vertex of the fragment to the plane e_i is calculated. All vertices are sorted by their distance to the plane. Then the nearest 1% of vertices are selected as candidates for the profile. For each of those vertices, all the patches to which they belong are filtered through a search in the patch list with their index number. In Figure 9 a sherd shaded by the value of distance to the axis of rotation is shown

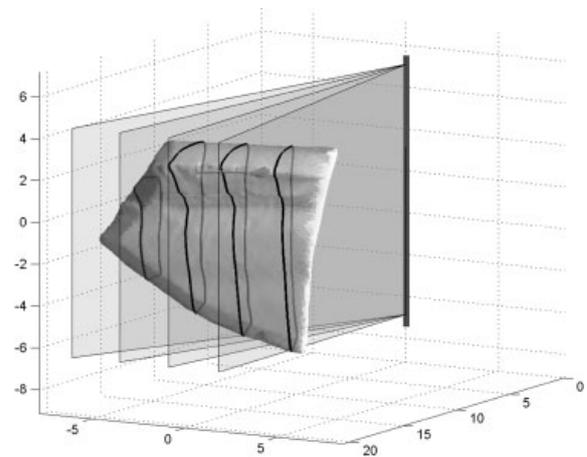


Figure 8. Sample of intersecting planes e_i .

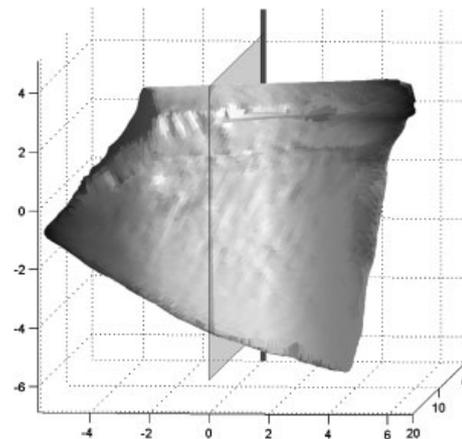


Figure 9. Properly oriented sherd and intersecting plane e_i . The grey values correspond to distances. Lighter values indicate closer distances to the intersecting plane.

(lighter values mean nearer to the intersecting plane). Every patch consists of three points that are connected through three lines. Every pair of vertices that has a point on each side of the plane is part of the profile line, because its connection intersects the plane. The coordinates of these pairs are rotated into the xy -plane and the z -coordinate is removed. The result is a properly oriented profile line.

- Next, the profile line with the longest elongation is computed: the difference between the maximum z -value and the minimum z -value of the profile line defines the height of the profile line. The remaining profile lines are used for evaluation of the estimation of the rotational axis.

Figure 10 shows two plots of diameters based on the profiles from two different fragments. The y -axis is the difference of the diameters to the overall mean diameter of all profiles in centimetres and the x -axis corresponds to the circular arc. The upper line shows the maximum diameter, the middle shows the mean diameter and the lower line shows the minimum diameter. The grey box visualizes the quality of the results by showing the overall mean diameter of all profiles versus the standard deviation. If the standard deviation exceeds a certain threshold (for example 0.5 cm) the fragment is excluded from further reconstruction.

Figure 10(a) shows a correct estimated rotational axis that results in mean diameter with a small standard deviation (smaller than 0.5 cm) along the perimeter of the sherd. Also the minimum and the maximum diameters are constant except on the left and right side, where the fracture of the fragment is located. In Figure 10(c) the mean diameter along the perimeter has a standard deviation of more than 0.5 cm (in this case

5 cm). This indicates that the estimation of the rotational axis is not accurate enough for further processing. In this case we plan to extend the algorithm for axis estimation by using additional information of the fragment, e.g. rills on the inner surface, detection of rim fragments.

Results

The resulting 3D reconstruction of fragments depends on the correct orientation of the profile section. The evaluation of the 3D representation is rather complicated since ground truth is not available due to the fact that there is no third dimension in archaeological archive drawings and the virtually reconstructed object does not exist in reality (only its fragment). The description of shape is subject to the ideas of the archaeologists and is not standardized.

In order to demonstrate the correctness of the computed profile lines, Figure 11 shows a recorded sherd (dark object) and its computed profile section (vertical line). The computation of the virtual fragment (grey object) is based on the profile section. One can see that the recorded fragment fits into the virtual fragment, which indicates that the computation is correct. Following multiple cross-sections along the perimeter of the virtual fragment (Figure 12a) one can observe hardly any deviation from the original fragment. Some are caused by the bumpiness of the surface, because the surface is not exactly rotationally symmetric, since it is hand-made pottery.

If the orientation of the fragment is incorrect, it does not fit into the virtual object and multiple cross-sections along the perimeter of the virtual fragment show large deviations from the original object (see Figure 12b).

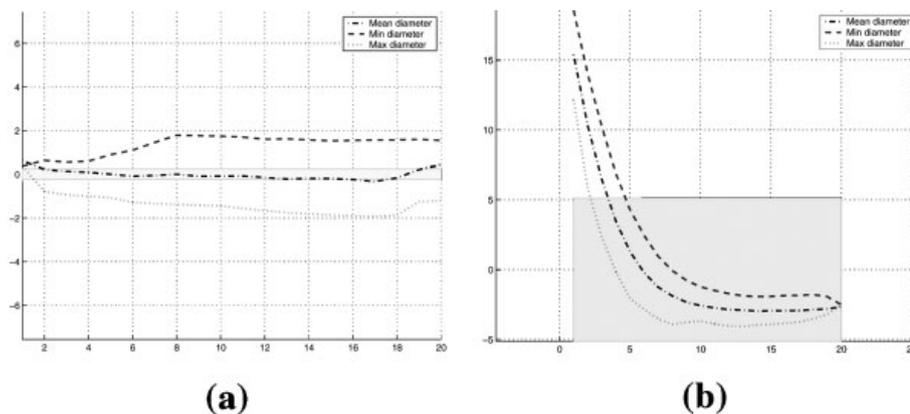


Figure 10. Maximum, mean and minimum diameters.

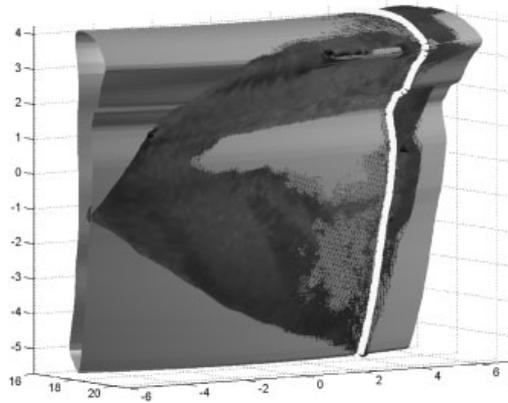


Figure 11. Reconstructed fragment, profile section and recorded fragment.

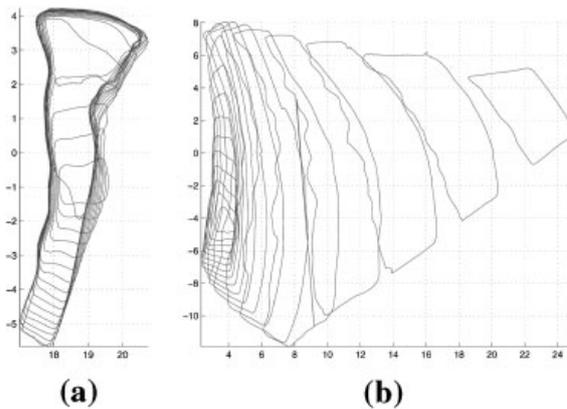


Figure 12. Multiple profile lines: (a) using a correctly estimated rotational axis; (b) using an incorrect rotational axis.

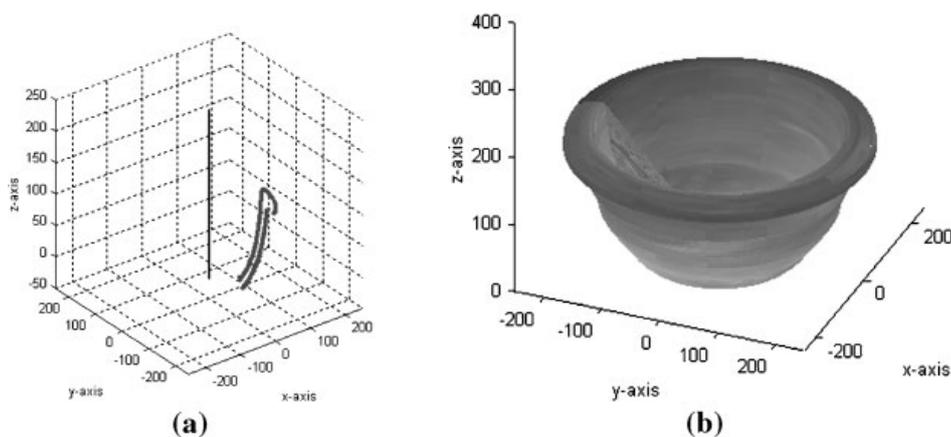


Figure 13. (a) 3D profile section; (b) 3D reconstruction from the profile section.

A computed profile and the axis of rotation are shown in Figure 13(a). It was rotated 360 degrees around the axis of rotation in order to construct the vessel in 3D. Next the resulting 3D point cloud was triangulated²⁹ and the acquired texture was mapped onto the triangulated mesh. Figure 13(b) shows the reconstructed pot.

In Figure 14 the reconstruction of the virtual pot was based on an incorrectly orientated rotational axis. Since the fragment is a base fragment (fragment from the bottom of the pot), has a flat surface with hardly any curvature it, and therefore it is not possible to orientate it by the use of its axis of rotation. The image is shown to demonstrate the deviation between the cross-section of an incorrect reconstruction and the original data (Figure 12b).

Figure 15 displays a reconstructed pot (grey object) out of a rim fragment (dark object) based on the profile line (light line) and its axis of rotation (dashed line).

Experiments were performed on 40 fragments of our pottery database. The success rate for correct extraction of the profile line and consequently the correct virtual reconstruction are around 50%. This has to be seen with respect to manual archival work done by archaeologists:³ for coarse ware around 35%³⁰ and for fine ware around 50%³¹ of the findings are used for further classification. It depends heavily on the shape of the fragment (e.g. handle, flat fragments like bottom pieces, small size). Eighteen fragments have been excluded from reconstruction due to wrong estimation of the axis of rotation.

Table 1 summarizes the results for 22 properly orientated fragments. Box and piece numbers are used for identification of the fragment. The radius r is the

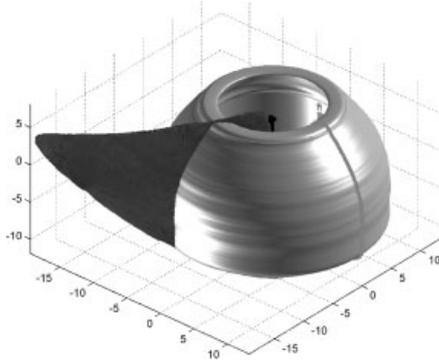


Figure 14. Incorrect reconstruction of a fragment due to wrong orientation.

estimated mean radius of the whole object. The standard deviation of the radius was estimated along the perimeter of the fragment. The thickness of the fragment is the difference between the mean radius of the inner side and the outer side. The fragment size is the percentage of the perimeter of the sherd compared to the perimeter of the whole object.

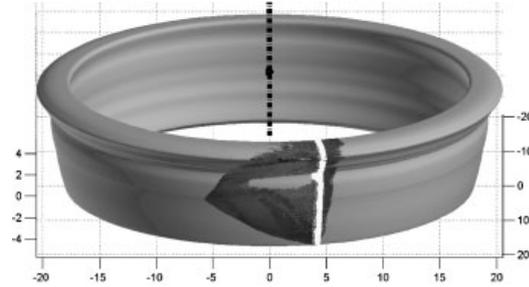


Figure 15. Reconstructed pot.

Experiments with synthetic data have shown that the correctness of the reconstruction depends on the correct estimation of the axis of rotation (see reference 28 for a detailed survey) and on the resolution of the 3D scanner used. The number of vertices of the data used ranges from 4000 to 15000, leading to a profile line with 200–300 points. The execution time using a prototype written in Matlab running on a Pentium III 1 GHz computer is less than a minute per sherd. It depends heavily on the computation of the axis of rotation (70–80% of the execution time).

Box no.	Sherd no.	Radius (cm)	Standard deviation (cm)	Mean thickness (cm)	Fragment size (%)
1	04	10.20	0.05	0.51	11.13
1	08	15.78	0.04	0.39	4.78
1	16	14.56	0.09	1.75	6.36
1	17	16.15	0.15	1.51	8.08
1	18	15.16	0.03	1.75	5.12
1	19	14.54	0.15	0.84	8.25
1	20	12.99	0.08	1.56	7.55
1	22	11.53	0.10	0.75	13.55
1	23	12.33	0.08	0.65	8.72
2	01	9.97	0.09	0.82	11.15
2	02	15.95	0.03	1.13	6.35
2	04	6.66	0.18	1.67	17.03
2	05	9.94	0.08	0.49	8.94
2	06	2.35	0.21	1.03	31.51
2	09	12.30	0.08	0.97	9.06
2	10	18.33	0.06	1.00	5.91
2	11	10.20	0.05	2.24	9.09
2	12	12.34	0.12	0.98	10.04
2	14	16.91	0.17	1.52	12.61
2	15	15.42	0.07	1.21	8.40
2	16	14.36	0.12	1.53	8.47
2	18	10.80	0.06	1.93	9.69

Table I. Results for 22 fragments

Conclusion and Outlook

We have proposed a prototype system for the automatic archival of archaeological fragments. The work was performed in the framework of the documentation of ceramic fragments. The methods proposed have been tested on synthetic and real data with reasonably good results since they are better than traditional manual archaeological radius and volume estimation. The ceramic documentation and reconstruction system described is currently under development to be integrated in the virtual excavation reconstruction project 3D-MURALE.³² Currently we are testing our reconstruction method on a larger test set taken from our 3D-MURALE test excavation site Sagalassos and test objects with ground truth (like fragments of industrially manufactured flower pots which are perfectly symmetric) to further evaluate our method. Next we will work on the classification system based on the profile in order to classify all profiles and to find matching fragments from similar vessels and finally from singular vessels.

ACKNOWLEDGEMENTS

The authors would like to thank Kristina Adler and Martin Penz from Vienna University, Institute of Classical Archaeology, for their archaeological support and contributions in evaluating the model representation approach and for helpful and inspiring discussions. This work was supported in part by the Austrian Science Foundation (FWF) under grant P13385-INF, the European Union under grant IST-1999-20273 and the Austrian Federal Ministry of Education, Science and Culture.

References

1. Fowler P, Boniface P. *Heritage and Tourism in 'The Global Village'*. Routledge: London, 1993.
2. Woodruff A, Aoki P, Hurst A, Szymanski M. Electronic guidebooks and visitor attention. In *Proceedings of the International Conference on Cultural Heritage and Technologies in the Third Millennium*, Vol. 1, Bearman D, Garzott F (eds). Milan, 2001; pp 437–454.
3. Orton C, Tyers P, Vince A. *Pottery in Archaeology*. Cambridge University Press: Cambridge, 1993.
4. Waldhaeusl P, Kraus K. Photogrammetrie für die Archäologie. In *Lebendige Altertumswissenschaft: Festgabe zur Vollendung des 70. Lebensjahres von Hermann Vetters dargebracht von Freunden, Schülern und Kollegen*, Kandler M (ed.). Vienna, 1985; pp 423–427.
5. Cooper D, et al. Assembling virtual pots from 3D measurements of their fragments. In *Proceedings of the International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, Vol. 1, Arnold D (ed.). Athens, 2001; pp 259–271.
6. Papaioannou G, Karabassi EA, Theoharis T. Virtual archaeologist: assembling the past. *IEEE Computer and Graphics and Applications* 2001; **21**(2): 53–59.
7. Copper D, et al. Bayesian pot-assembly from fragments as problems in perceptual-grouping and geometric-learning. In *Proceedings of 16th International Conference on Pattern Recognition*, Vol. 1, Kasturi R, Laurendeau D, Suen C (eds), Quebec City, IEEE Computer Society, 2002; pp 297–302.
8. Kong W, Kimia BB. On solving 2D and 3D puzzles using curve matching. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, 2001; pp 583–590.
9. Leitao HCG, Stolfi J. A multiscale method for the reassembly of two-dimensional fragmented objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2002; **24**(9): 1239–1251.
10. Kappel M, Sablatnig R. Automated 3D recording of archaeological pottery. In *Proceedings of the International Conference on Cultural Heritage and Technologies in the Third Millennium*, Vol. 1, Bearman D, Garzott F (eds). Milan, 2001; pp 169–182.
11. Sablatnig R, Menard C. 3D reconstruction of archaeological pottery using profile primitives. In *Proceedings of the International Workshop on Synthetic-Natural Hybrid Coding and Three-Dimensional Imaging*, Sarris N, Strintzis MG (eds). Rhodes, Greece, September 1997; pp 93–96.
12. Rosenfeld A, Nakamura A. Local deformations of digital curves. *Pattern Recognition Letters* 1997; **18**(7): 613–620.
13. Shoukry A, Amin A. Topological and statistical analysis of line drawings. *Pattern Recognition Letters* 1983; **1**: 365–374.
14. Kappel M, Sablatnig R. Automated segmentation of archaeological profiles for classification. In *Proceedings of 16th International Conference on Pattern Recognition*, Vol. 1, Kasturi R, Laurendeau D, Suen C (eds). IEEE Computer Society: Quebec City, 2002; 57–60.
15. Kandler M. The legionary fortress of Carnuntum. *Anzeiger der österreichischen Akademie der Wissenschaften*. 1979; **115**: 335–351.
16. Grünwald M. Ausgrabungen im Legionslager von Carnuntum (Grabungen 1969–1977). *Keramik und Kleinfunde 1976–1977*. In *Der Römische Limes in Österreich 34*, Vienna, 1986; pp 10–11.
17. DePiero F, Trivedi M. 3-D computer vision using structured light: design, calibration, and implementation issues. *Advances in Computers* 1996; **43**: 243–278.
18. Klette R, Koschan A, Schlüns K. *Computer Vision: Räumliche Information aus digitalen Bildern*. Vieweg: Germany, 1996.
19. Cosmas J, et al. 3D MURALE: a multimedia system for archaeology. In *Proceedings of International EuroConference on Virtual Reality, Archaeology and Cultural Heritage*, Athens, Greece, 2001.
20. Ben Yacoub S, Menard C. Robust axis determination for rotational symmetric objects out of range data. In *21st Workshop of the Oeagm*, Burger W, Burge M (eds). Hallstatt, Austria, May 1997; pp 197–202.
21. Pottmann H, Peternell M, Ravani B. An introduction to line geometry with applications. *Computer-Aided Design* 1999; **31**: 3–13.

22. Oja E, Xu L, Suen CY. Modified Hebian learning for curve and surface fitting. *Neural Networks* 1992; 5(3): 441–457.
23. Oja E. *Subspace Methods of Pattern Recognition*. Wiley: Chichester, 1983.
24. Bhanu B. Representation and shape matching of 3D objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984; 6(3): 340–351.
25. Greenspan M, Godin G. A nearest neighbor method for efficient ICP. In *Proceedings of the 3rd IEEE International Conference on 3-D Digital Imaging and Modeling*, Quebec, 2001; pp 161–168.
26. Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. In *Proceedings of the 3rd IEEE International Conference on 3-D Digital Imaging and Modeling*, Quebec, 2001; pp 145–152.
27. Pulli K. Multiview registration for large data sets. In *Proceedings of the 2nd IEEE International Conference on 3-D Digital Imaging and Modeling*, Ottawa, 1999; pp 160–168.
28. Sablatnig R, Kampel M. Model-based registration of front- and backviews. *Computer Vision and Image Understanding* 2002; 87(1–3): 90–103.
29. Klette R, Rosenfeld A, Sloboda F. *Advances in Digital and Computational Geometry*. Springer: Singapore, 1998.
30. Degeest R. The common wares of Sagalassos. In *Studies in Eastern Mediterranean Archaeology*, Walkens M (ed.). Brepols: Turnhout, 2000.
31. Poblome J. Sagalassos red slip ware. In *Studies in Eastern Mediterranean Archaeology*, Walkens M (ed.). Brepols: Turnhout, 1999.
32. Sablatnig R, Kampel M. The virtual reconstruction of an archaeological site: an overview of the MURALE project. In *Proceedings of the 6th Computer Vision Winter Workshop 2001*, Likar B (ed.). Bled, Slovenia, Slovenian Pattern Recognition Society, February 2001; pp 60–70.

Authors' biographies:



Martin Kampel was born in Steyr, Austria, in 1968. He studied data technologies (1990–1993) and computer science (1993–1999) at the Vienna University of Technology. Since 1996 he is working at the Pattern Recognition and Image Processing Group, Institute of Computer-Aided Automation in Vienna, engaged in research, teaching and administration. His research interests are 3D-Vision, Computer Graphics and Virtual Archaeology. Currently he is working on his PhD thesis on 3D mosaicing.



Robert Sablatnig was born in Klagenfurt, Carinthia, Austria, in 1965. He received the B.Sc. degree in Computer Science in 1988, the M.Sc. degree (Diplom Ingenieur) in Computer Science (Computer Graphics, Pattern Recognition & Image Processing) in 1992 and the Ph.D. degree in Computer Science in 1997 from the Vienna University of Technology.

He is an assistant professor (Univ.Ass.) of computer vision at the Pattern Recognition and Image Processing Group, Institute of Automation, engaged in research, project leading, industry consulting and teaching. His research interests are Applications in Industrial Inspection, Automatic Visual Inspection, KB- Visual Inspection Systems, 3d- Machine Vision and Hierarchical Pattern Recognition. He is author or co-author of more than 80 scientific publications presented at several international conferences and workshops and is a member of the IAPR and the IEEE.

Chapter 4

On 3D Mosaicing of Rotationally Symmetric Ceramic Fragments

Martin Kampel and Robert Sablatnig. On 3D Mosaicing of Rotationally Symmetric Ceramic Fragments, In J. Kittler, M. Petrou, and M. Nixon, editors, *Proceedings of 17th International Conference on Pattern Recognition*, IEEE Computer Society, Vol. 2, pp. 265-268, 2004.

On 3D Mosaicing of Rotationally Symmetric Ceramic Fragments

Martin Kampel and Robert Sablatnig
Vienna University of Technology
Institute of Computer Aided Automation
Pattern Recognition and Image Processing Group
Favoritenstr.9/183-2, A-1040 Vienna, Austria
{kempel,sab}@prip.tuwien.ac.at

Abstract

A major obstacle to the wider use of 3D object reconstruction and modeling is the extent of manual intervention needed. Such interventions are currently massive and exist throughout every phase of a 3D reconstruction project: collection of images, image management, establishment of sensor position and image orientation, extracting the geometric detail describing an object, merging geometric, texture and semantic data. This work aims to develop a solution for automated documentation of archaeological pottery, which also leads to a more complete 3D model out of multiple fragments. Generally the 3D reconstruction of arbitrary objects from their fragments can be regarded as a 3D puzzle. In order to solve it we identified the following main tasks: 3D data acquisition, orientation of the object, classification of the object and reconstruction. We demonstrate the method and give results on synthetic and real data.

1 Introduction

Reassembly of fragmented objects from a collection of thousands randomly mixed fragments is a problem that arises in several applied disciplines, such as archaeology, failure analysis, paleontology, art conservation, and so on. Solving such jigsaw puzzles by hand may require years of tedious and delicate work, consequently the need for computer aided methods is obvious [8]. The assembling of an object from pieces is called mosaicing [7]. It is similar to the automatic assembly of jigsaw puzzles [1]. In archaeology, most of the finds are in form of fragments, especially in the area of ceramics. Therefore mosaicing is of great interest in this field, since it enables both, a real and a virtual reconstruction of the original object. Most of the ceramic is rotationally symmetric since it was produced on a potter's wheel. Using this fact, one can solve the mosaicing problem even if there are gaps between the fragments, just like a human would solve this problem. Figure 1a shows a box

filled with archaeological fragments, which possibly could fit to each other. Figure 1b illustrates manually identified, matching fragments.

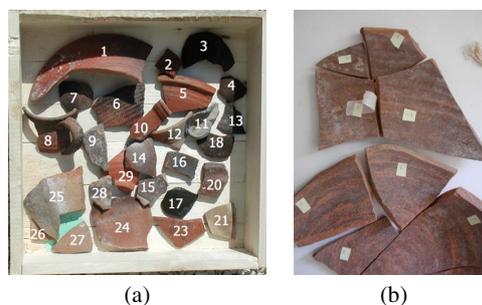


Figure 1. Archaeological objects: (a) Box with possibly, matching fragments, (b) Matching fragments.

More generally mosaicing can be seen as a special case of object recognition by approximate outline matching: The specific problem of identifying adjacent ceramic fragments by matching the shapes of their outlines was considered by Üçoluk and Toroslu [13], Hori et. al. [2], Kong et. al. [5] and Kanoh et. al. [4]. Marques et al. [9] present a 2D object matching technique based on the comparison of a reference contour to the contours in the image partition. Similarly, Leitão and Stolfi [8] demonstrate a multi-scale matching method and Papaioannou et. al. present a semi-automatic reconstruction of archaeological finds [10]. We observe a main focus on the analysis of the outline of the break curve: 2D outline matching is most common [8, 4, 5, 1, 6], but work on 3D outline matching exist [13]. Surface matching of fractured surfaces is proposed in [10]. So far, no complete system from acquisition to reconstruction has been described.

This paper focuses on the reconstruction of pottery out of many fragments based on the profile. With respect to

our previous work [12], the paper describes the finding and matching of candidate fragments as its main contribution. Our approach to pottery reconstruction is based on the following main tasks: After acquiring 3D data with the Minolta VIVID 900, we start with the estimation of the correct orientation of the fragment, which leads to the exact position of a fragment on the original vessel. Next, the classification of the fragment based on its profile section allows us to decide to which class an object belongs to, presented in Section 2. Since we know the orientation of the candidate fragments we defined a two-degrees-of-freedom search space for representing the alignment of two fragments. A matching algorithm based on the point-by-point distance between facing outlines is proposed in Section 3. Reconstruction results on synthetic and real data are given in Section 4, followed by conclusions and outlook on future work.

2 Determination of matching candidates

Archaeological pottery is assumed to be rotationally symmetric since it was made on a rotation plate. With respect to this property the axis of rotation is calculated using a Hough inspired method [12].

In order to reconstruct complete pots out of fragments, profiles with similar attributes are to be found in an archive database. Classification of newly found fragments of unknown type is performed by comparing the description of the new fragment with the description of already classified fragments. The fragment structure is formed by its *shape features* (or geometric features like the profile) and its *properties* (or material like clay, color and surface). The description of the fragment is structured in a description language consisting of primitives and relations. Primitives are a representation of shape features, relations represent the properties.

The description language, which was originally designed to solve 2D automatic visual inspection problems [11], is applied and extended in order to solve the classification problems. The actual profile contains features, which are a representation of shape features. To accomplish classification, primitives are further subdivided into part-models (or part-primitives), the consistency between part-primitives is established by relations among part parameters (see [11] for details).

This method enables us to compute the confidence of a node by summing up the weighted tolerances of each attribute of the node and the overall confidence of the sub-graph connected to this node. By computing the consistency for different descriptions, the one with the highest confidence value can be chosen if the confidence is above a certain threshold. For a given profile all primitives are represented in the description of the profile.

3 Fragment Matching

The optimal pairing of matching candidates obtained serves as input for the fragment matching part. Consequently we know those pairs of fragments which were probably adjacent in the original object. In order to represent the matching of two fragments, G. Papaioannou et. al. [10] describe seven pose parameters. In their approach the two fragments are first prealigned so that their broken facets face each other. In our case we know the orientation of a fragment, consequently we prealign two candidate fragments by simply aligning their axis of rotation. As a result, a two-degrees-of-freedom continuous search space is defined. The transformation which matches two candidate fragments consists of a translation along the z-axis with parameter T_z and a rotation around the z-axis with parameter R_z (see Figure 2).

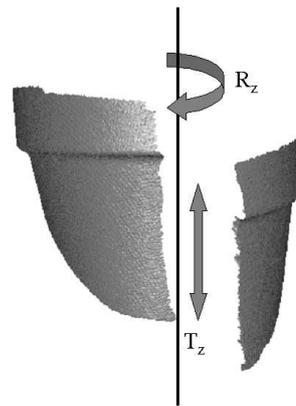


Figure 2. Fragment Matching with 2-degrees-of-freedom.

The basic concept in our method for estimating R_z is that the best fit is likely to occur at the relative pose which minimizes the point-by-point distance between the facing outlines. For this reason, we introduce a matching error ϵ_M based on the mean Euclidian distance between the corresponding points of the outlines of the candidate fragments with points $X = (x, y)$ and $X' = (x', y')$:

$$\epsilon_M = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2}. \quad (1)$$

where N is the number of data points used. The height of the fragment, limits the length of the matching segments. Different fragments types lead to the following matching possibilities:

A Rim fragments: first T_z is computed by aligning the rim along the orifice plane [3]. Next R_z is estimated,

so that the positioning transformation with the smallest matching error ϵ_M is considered to be the correct position.

- B Bottom fragments: first T_z is computed by aligning the bottom along the base plane. Next R_z is estimated in the same way as for rim fragments.
- C Wall fragments: Candidates are first aligned along their profile sections. Next R_z is estimated in the same way as for rim fragments. Since it is not clear whether a new candidate fragment is in bottom up or bottom down position, we have to compute R_z and T_z for both positions. The positioning transformation with the smallest matching error ϵ_M is considered to be the correct position.

Matching algorithm

1. Define reference fragment F_{ref} with its axis of rotation ROT_{ref} : defines a new pot P , creates the pot coordinate system, ROT_{ref} is aligned to the z-axis.
2. Prealignment of the candidate fragment F_{cand} by its axis of rotation ROT_{cand} : ROT_{cand} is aligned to ROT_{ref} . This results in a two-degrees-of-freedom search space: Translation T_z along the axis of rotation (up/down) and rotation R_z around the axis of rotation.
3. Estimation of the translation parameter T_z : search for minimal distance d between all y-values (radius) of the profile of F_{ref} and the profile of F_{cand} . Exception A: Rim fragments are aligned along the orifice plane. Exception B: Bottom fragments are aligned along the base plane. When the candidate fragment is a wall fragment, the minimal distance d is computed for both positions, and the one with the smaller is considered to be the correct position.
4. Estimation of the rotation parameter R_z by finding the position with the smallest matching error ϵ_M .

4 Results

In order to evaluate the results we have tested our method on synthetic 3D data of three parts of a synthetic pot. The orientation of the fragments is defined, which leads to three perfect matching parts. The experiment has shown a 100% theoretical accuracy of the approach.

In order to get data of matching fragments of a whole pot, we broke a flowerpot into 5 parts. We got three rim fragments, one wall fragment and one bottom fragment. Each part was digitized leading to a front and back view of each fragment. The biggest part (nr. 2) covers half of the pot and consists of 135070 triangles, whereas the smallest consists of 8210 triangles. Next we computed the orientation of the

fragments, which leads to four matching candidates and one not processable object: a large part of the bottom fragment (Part 4) consists of flat area. It was therefore excluded from further processing due to its curvature being too low.

Starting with part one as reference fragment for each candidate a matching error was computed. Next part two was defined as reference fragment and again for each remaining candidate a matching error was computed. This procedure was continued until no candidate remained. Table 1 summarizes T_z , R_z and the matching errors for each possible candidate. $RF_{nr.}$ and $CF_{nr.}$ denote the number of the reference fragment and the number of candidate fragment respectively, and ϵ_M denotes the matching error. The value of ϵ_M for correct matches ranges from 1.12 to 0.63, the combination of part 3 and 5 shows an incorrect match with an error ϵ_M of 12.92.

Figure 3a displays the resulting match of part 1 and part 3 as both parts are rim fragments. Figure 3b shows the resulting match of part 1 and part 5. Since part 5 is a wall fragment the ϵ_M was computed for both possible positions, and the position with lower ϵ_M was finally chosen.

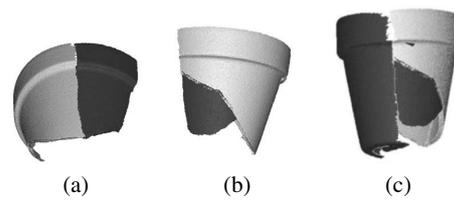


Figure 3. Matched parts: (a) part 1 and part 3 (b) part 1 and part 5 (c) Matching parts 1, 2, 3, and 5.

Figure 3c shows the final reconstruction of the pot. Correct matches for all four candidate fragments have been found. The missing bottom of the pot is due to part 4, not being processable because of its flat shape.

We applied our technique to real archaeological fragments (Nr: 319-71, 209-71 from the late Roman burnished ware of Carnuntum [12]). Both pieces are rim fragments (Figure 4a and b). The alignment along the orifice plane allowed the estimation of $T_z = 7.49\text{cm}$. The smallest $\epsilon_M = 0.31$ was found for $R_z = 3.35^\circ$. Figure 4c shows the matched outlines of the two fragments and Figure 4d shows the final reconstruction.

Another example on real archaeological fragments was done on the common ware of Sagalassos [3]. One rim and two wall fragments were recorded and processed. Table 1 summarizes T_z , R_z and the matching errors for each possible candidate. Correct matches were found between part one and part two ($\epsilon_M = 1.32$) and part two and part three ($\epsilon_M = 1.21$). No correct match was found between part one and part three ($\epsilon_M = 14.81$), because there was no

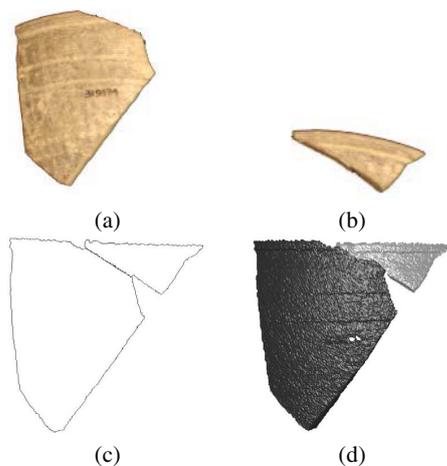


Figure 4. Archaeological rim fragments: (a) Part 1, (b) Part 2, (c) Matching outlines, (d) Matching parts.

alignment of the profile sections (part one is on top of part three). Nevertheless all three fragments were matched together, since the matching of part two succeeded for both candidates.

Ware	$RF_{n.r.}$	$CF_{n.r.}$	T_z [mm]	R_z [deg]	ϵ_M
Flowerpot	1	2	12.03	22.81	1.12
Flowerpot	1	3	8.67	-41.29	0.81
Flowerpot	1	5	9.34	73.21	0.63
Flowerpot	2	3	-4.94	17.61	0.92
Flowerpot	2	5	-10.02	-26.75	0.71
Flowerpot	3	5	11.10	32.99	12.92
Carnuntum	1	2	7.49	3.35	0.31
Sagalassos	1	2	-4.29	11.70	1.32
Sagalassos	1	3	-1.61	7.59	14.81
Sagalassos	2	3	-5.19	15.76	1.21

Table 1. Results of the matching process.

The results demonstrate the possibility of automatically matching adjacent fragments by our method. It works for fragments which can be oriented and classified by our approach with one exception: two adjacent fragments on top of each other cannot be matched by our method, because they do not have overlapping profile sections. Furthermore if the surface of the fragment is too flat or too small or the classification is not known, the fragment is not considered for reconstruction.

5 Conclusion

We have proposed a method for the assembly of an object from pieces, which in our case means the reconstruction of an archaeological pot from its fragments. The outcome on vessel reconstruction out of multiple fragments was described by real 3D data. The ceramic documentation and reconstruction system described was recently integrated into the virtual excavation reconstruction project 3D MURALE. Future work will be directed towards setting up a pottery database with more than 100 fragments and applying the algorithm to find matching pieces.

References

- [1] G. C. Burdea and H. J. Wolfson. Solving Jigsaw Puzzles by a Robot. *IEEE Transactions on Robotics and Automation*, 5(5):752–764, 1989.
- [2] K. Hori, M. Imai, and T. Ogasawara. Joint Detection for Potsherds of Broken Earthenware. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, volume 2, pages 440–445, June 1999.
- [3] M. Kampel. *3D Mosaicing of Fractured Surfaces*. PhD thesis, TU-Vienna, Inst. f. Automation, Pattern Recognition and Image Processing Group, 2003.
- [4] M. Kanoh1, S. Kato, and H. Itoh. Earthenware reconstruction based on the shape similarity among potsherds. *Society for Science on Form*, 16(1):77–90, 2001.
- [5] W. Kong and B. Kimia. On solving 2D and 3D puzzles using curve matching. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, volume 2, pages 583–590, 2001.
- [6] D. Kosiba, P. M. Devaux, S. Balasubramanian, T. Gandhi, and R. Kasturi. An Automatic Jigsaw Puzzle Solver. In *Proc. of the 12th IAPR International Conference on Pattern Recognition, Jerusalem*, volume 1, pages 616–618. IEEE-Computer Society Press, 1994.
- [7] Langenscheidt. *Langenscheidts Fremdwörterbuch*. Friedhelm Hübner, 1989.
- [8] H. C. G. Leitão and J. Stolfi. A Multiscale Method for the Reassembly of Two-Dimensional Fragmented Objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(9):1239–1251, 2002.
- [9] F. Marques, M. Pardis, and R. Morros. Object matching based on partition information. In *IEEE Computer Society International Conference on Image Processing*, volume 2, pages 829–832, 2002.
- [10] G. Papaioannou, E. Karabassi, and T. Theoharis. Virtual archaeologist: Assembling the past. *IEEE Computer Graphics*, 21(2):53–59, March-April 2001.
- [11] R. Sablatnig. *A Highly Adaptable Concept for Visual Inspection*. PhD thesis, TU-Vienna, Inst. f. Automation, Pattern Recognition and Image Processing Group, 1997.
- [12] R. Sablatnig and M. Kampel. Model-based registration of front- and backviews. *Computer Vision and Image Understanding*, 87(1):90–103, 2002.
- [13] G. Üçoluk and H. Toroslu. Automatic Reconstruction of Broken 3D surface objects. *Computers and Graphics*, 23(4):573–582, 1999.

Chapter 5

Rule based System for Archaeological Pottery Classification

Martin Kampel and Robert Sablatnig. Rule based System for Archaeological Pottery classification, *Pattern Recognition Letters*, Elsevier Science Inc., 28(6):740-747, 2007.



Rule based system for archaeological pottery classification [☆]

Martin Kampel ^{*}, Robert Sablatnig

*Vienna University of Technology, Institute of Computer Aided Automation, Pattern Recognition and Image Processing Group 183/2,
Favoritenstr. 9, A-1040 Wien, Austria*

Available online 5 October 2006

Abstract

Motivated by the requirements of the present archaeology, we are developing an automated system for archaeological classification of ceramics. Classification and reconstruction of archaeological fragments is based on the profile, which is the cross-section of the fragment in the direction of the rotational axis of symmetry. In order to segment the profile into primitives like rim, wall, and base, rules based on human expert knowledge are created. The input data for the estimation of the profile is a set of points produced by the acquisition system. A function fitting this set is constructed and later on processed to find the characteristic points necessary to classify the original fragment. We demonstrate the method and give results on real archaeological data.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Archaeological ceramics; Typology and classification; Curvature analysis; 3D acquisition

1. Introduction

Ceramics are one of the most widespread archaeological finds and are a short-lived material. This property helps researchers to document changes of style and ornaments. Therefore, ceramics are used to distinguish between chronological and ethnic groups. Furthermore ceramics are used in the economic history to show trading routes and cultural relationships. Especially ceramic vessels, where shape and decoration are exposed to constantly changing fashion, not only allow a basis for dating the archaeological strata, but also provide evidence of local production and trade relations of a community as well as the consumer behavior of the local population (Orton et al., 1993).

In order to make a later classification possible, the object is described in different ways: shape, decoration,

technological manufacturing stage, material, and color. Shape is usually described in terms of type-series using traditional pottery classification systems. For this, the description of the profile is important. Decoration, material, color, and a careful examination of the traces left on the vessel which indicate the steps taken during the manufacturing process is a further important property to be investigated in order to perform a subsequent classification.

All descriptions of the object serve only one final aim: the correct *classification* of the material. There are several different ways of classifying vessels: based on their shape, rim form, the presence of handles or spouts, decorative motifs and so on. The purpose of classification is not only to get a systematic view of the material found but also to find different fragments belonging to the same vessel based on attributes stored in an archive database. After that, the profile of the fragment can be used to reconstruct the original (complete) vessel. This includes the possibility of reconstructing missing parts of the vessel and the search for possible matches of other fragments already stored in the archive with the one that is under consideration (part-assembly).

Up to now all this has been done manually which means a lot of routine work for archaeologists and a very

[☆] This work was partly supported by the Austrian Science Foundation (FWF) under grant P13385-INF, the European Union under grant IST-1999-20273 and Culture and by the innovative project ‘3D technology’ of the Vienna University of Technology.

^{*} Corresponding author. Tel.: +43 15880118364.

E-mail addresses: kampel@prip.tuwien.ac.at (M. Kampel), sab@prip.tuwien.ac.at (R. Sablatnig).

URL: www.prip.tuwien.ac.at (M. Kampel).

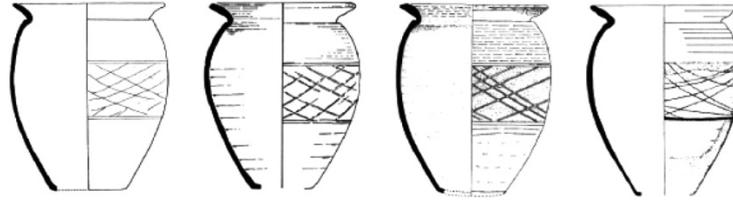


Fig. 1. The limits of objective recording: the same vessel drawn by four different illustrators (from Orton et al., 1993).

inconsistent representation of the real object (Orton et al., 1993). First, there may be errors in the measuring process, diameter or height may be inaccurate, second, the drawing of the fragment should be in a consistent style, which is not possible since a drawing of an object without interpreting it is very hard to do. The third process, grouping or classifying, is also a very difficult task.

There have been several attempts to develop a reliable method of classification (Binford, 1965; Adams and Adams, 1991; Sinopoli, 1991), none of which is widely accepted. A graphic documentation done by hand additionally raises the possibility of errors. This leads to a lack of objectivity in the documentation of the material found. To give an example, a vessel was drawn by four different illustrators resulting in four different vessels as shown in Fig. 1. Note the different shape and decoration, the rim and the thickness for instance are significantly different.

Gilboa et al. (2004) presents a strategy that pertains to typology and classification which are confined to shape attributes. Their analysis is based on the description of a planar curve in terms of its curvature function. Their main source of error is that the information is derived from hand drawn profiles, whose accuracy cannot be assessed. Here we use digitized information for extracting quantitative measures for various measures of the profile.

This paper is organized as follows. In Section 2, we describe the acquisition method for 3D-shape estimation and how we estimate the profile sections. Next we show the determination of shape characteristics based on so-called characteristic points in Section 3. The generation of primitives is presented in Section 4. Results using real archaeological data are given in Section 5, followed by conclusions and outlook on future work.

2. Data acquisition

The acquisition method for estimating the 3D-shape of a fragment is shape from structured light (DePiero and Trivedi, 1996), which is based on active triangulation (Besl, 1988). We used the *Vivid 900* 3D Scanner developed by MINOLTA.¹

Fig. 2 illustrates the acquisition setup consisting of the Vivid 900 Scanner connected to a PC and the object to be recorded. Optionally the object is placed on a turntable

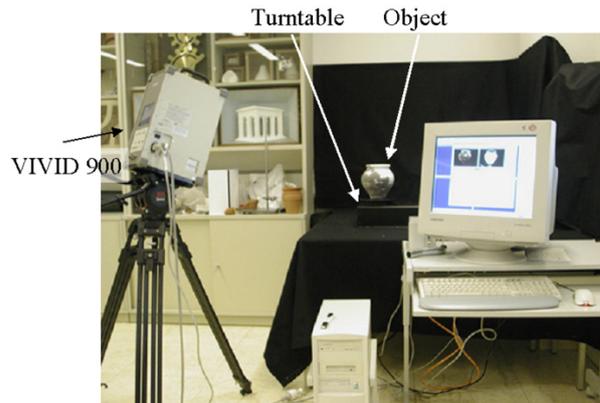


Fig. 2. The Minolta VIVID 900 Scanner.

with a diameter of 40 cm, whose desired position can be specified with an accuracy of 0.1° . The 3D Scanner works on the principle of laser triangulation combined with a color CCD image. It is based on a laser-stripe but a galvanometer mirror is used to scan the line over the object.

We have carried out in situ tests to capture 3D sherds and other finds from the excavation site CARNUNTUM (Gruenewald, 1986) in Austria. For each fragment two views (front- and back-views) have been recorded. Common sizes of fragments range from $10 \times 10 \times 8$ cm to $30 \times 30 \times 20$ cm, and they are recorded with an accuracy of 0.5 mm. The 3D-data is stored as a software independent *VRML*-file containing 3D-points that are connected in form of triangles. Our pottery dataset currently consists of 40 different sherds.

Archaeological pottery is assumed to be rotationally symmetric since it was made on a rotation plate. With respect to this property the axis of rotation is calculated using a Hough inspired method (Sablatnig and Kampel, 2002). To perform the registration of the two surfaces of one fragment, we use a priori information about fragments belonging to a complete vessel: both surfaces have the same axis of rotation since they belong to the same object. Consequently, we register the two views by calculating the axis of rotation and by bringing the resulting axes into alignment. A detailed description of the registration algorithm can be found in (Sablatnig and Kampel, 2002). The technique proposed works best for wheel-thrown pottery, and would not be suitable for hand-made pottery, since it relies on the rotational symmetry.

¹ MINOLTA Austria GMBH: www.minolta.com.

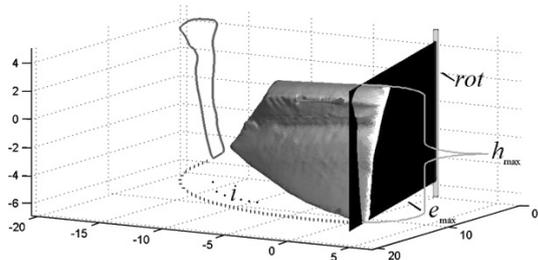


Fig. 3. Orientated sherd, rotational axis, intersecting plane e_{\max} and longest profile line.

The registration of front- and back-view together with the axis of rotation provide the profile used to reconstruct the vessel.

Fig. 3 shows the 3D-model of a sherd and its rotational axis rot as a vertical line along the z -axis. The black plane represents the intersecting plane e_{\max} at the maximum height h_{\max} of the sherd. The longest profile line is the longest line along the surface of the sherd parallel to the rotational axis rot . The radius r is the estimated mean radius of the profile line. The extracted profile line is shown in the xz -plane. Our algorithm for the estimation of the longest profile line consists of the following steps:

- (1) First the axis of rotation is transformed into the z -axis of the coordinate system in order to simplify further computation.
- (2) The fragment's size described by its circular arc is estimated. Depending on the size we compute a number of intersecting planes e_i , which are used for the profile estimation. The number of planes e_i depends on the length of the perimeter of the fragment. Experiments have shown that 7–13 profile lines return the best ratio of exactness and performance. Fig. 4 shows a sample of four planes e_i intersecting the 3D-model and the plots of the extracted profile lines on the surface of the sherd.

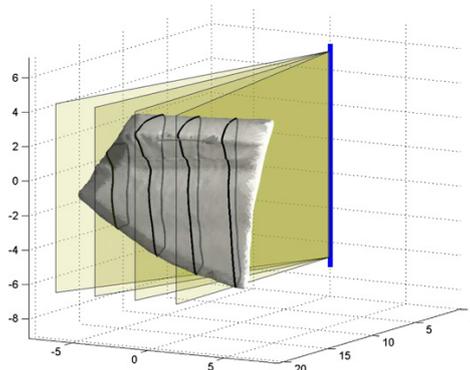


Fig. 4. Sample of intersecting planes e_i .

- (3) A profile line is calculated by intersecting the 3D-data of the fragment with planes e_i : first the distance of each vertex of the fragment to the plane e_i is calculated. All vertices are sorted by their distance to the plane. Then the nearest 1% of vertices are selected as candidates for the profile. The result is a properly oriented profile line.
- (4) Next the profile line having the longest line is found: the difference between the maximum z -value and the minimum z -value of the profile line defines the height of the profile line. The remaining profile lines are used for evaluation of the estimate of the rotational axis.

Fig. 5 shows two plots of diameters based on the profiles from two different fragments. The y -axis is the difference of the diameters to the overall mean diameter of all profiles in centimeters and the x -axis corresponds to the circular arc. The upper line shows the maximum diameter, the middle shows the mean diameter and the lower line shows the minimum diameter. The grey box allows the quality of the results to be visualized by showing the overall mean diameter of all profiles versus the standard deviation of all profiles. If the standard deviation exceeds a certain threshold

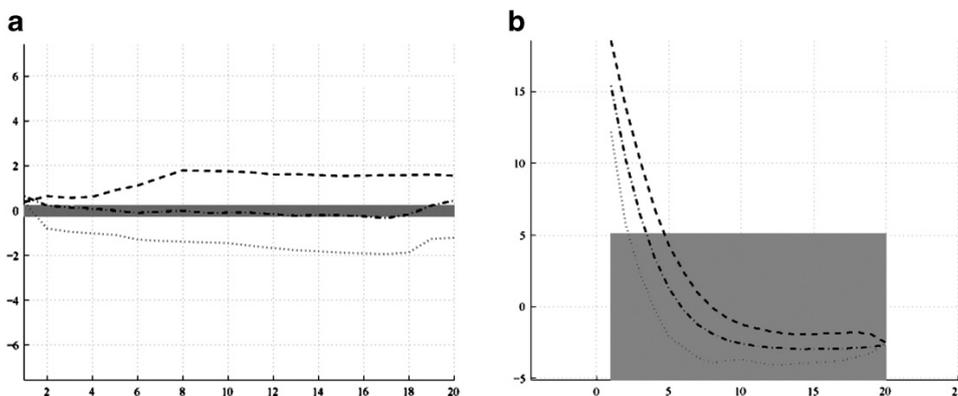


Fig. 5. Maximum, mean and minimum diameters.

(for example 0.5 cm) the fragment is excluded from further reconstruction.

As can be seen in Fig. 5a a correctly estimated rotational axis results in mean diameter with a small standard deviation (smaller than 0.5 cm) along the perimeter of the sherd. Also the minimum and the maximum diameter are constant except on the left and right side, where the fracture of the fragment is located. In Fig. 5b the mean diameter along the perimeter has a standard deviation of more than 0.5 cm (in this case 5 cm). This indicates that the estimate of the rotational axis is not accurate enough for further processing. In this case, we plan to extend the algorithm for axis estimation by using additional information on the fragment e.g. rills on the inner surface, detection of rim-fragments.

Experiments were done on all 40 fragments of our pottery database. The success rate for correct extraction of the profile line and consequently the percentage of sherds used for further reconstruction is around 50% of the data found at the excavation site. This should be compared to manual archivation done by archaeologists (Orton et al., 1993): for coarse ware around 35% (Degeest, 2000) and for fine ware around 50% (Poblome, 1999) of the findings are used for further classification. It depends heavily on the shape of the fragment (e.g. handle, flat fragments like bottom pieces, small size, etc.). Eighteen fragments have been excluded from reconstruction due to incorrect estimation of the axis of rotation. Details on the evaluation can be found in (Kampel and Sablatnig, 2003).

3. Determination of shape characteristics

By classifying the parts of the profile, the complete vessel is classified, and missing parts may be reconstructed. Following the manual strategy of the archaeologists, the profile should first be automatically segmented into its parts, the so-called *primitives* (Kampel and Sablatnig, 2002). Our approach to do so is a hierarchical segmentation of the profile into rim, wall, and base by creating segmentation rules based on expert knowledge of the archaeologists and the curvature of the profile. These three segments of the profile are stored in a so-called *description* of the profile. Fig. 6 shows an archive drawing of a frag-

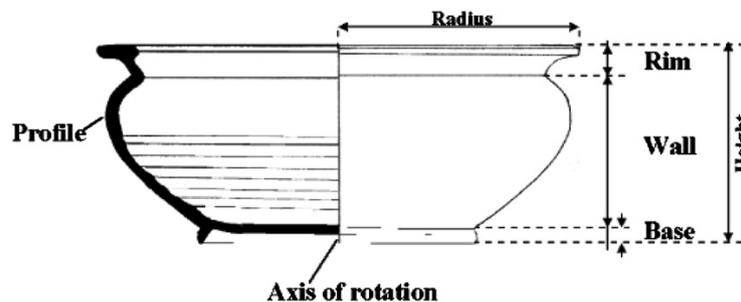


Fig. 6. Profile with known primitives.

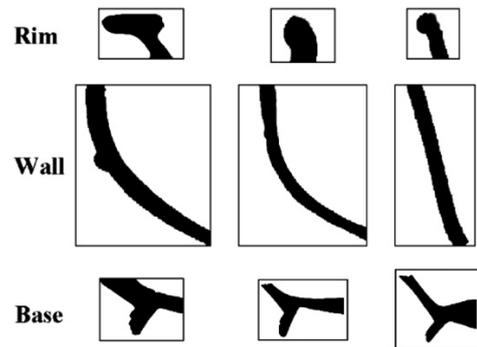


Fig. 7. Different shapes of primitives.

ment with its profile section divided into various primitives. If there is a corner point, that is a point at which the curvature changes “substantially”, the segmentation point is obvious. If there is no corner point, the segmentation point has to be determined mathematically (Adams and Adams, 1991).

Up to now this segmentation has been done manually by archaeologists, and there are no segmentation standards in archaeology (Adams and Adams, 1991). Fig. 7 shows different shapes of manually segmented primitives.

The basis for the segmentation is the outer profile line, i.e. the profile line along the outside of the vessel. The segments of the profile are divided by so-called characteristic points or segmentation points. Fig. 8 shows the classification scheme applied to an “S-shaped vessel” as an example. The coordinate system has its origin at the intersection point of the axis of rotation and the orifice line.

In order to allow proper segmentation, the following points have been identified.

- SP, starting point: in the case of vessels with a horizontal rim: innermost point, where the profile line touches the orifice plane.
- OP, orifice point: outermost point, where the profile line touches the orifice plane.
- IP, inflexion point: point, where the curvature changes its sign, i.e. where the curve changes from a left turn to a right turn or vice versa.

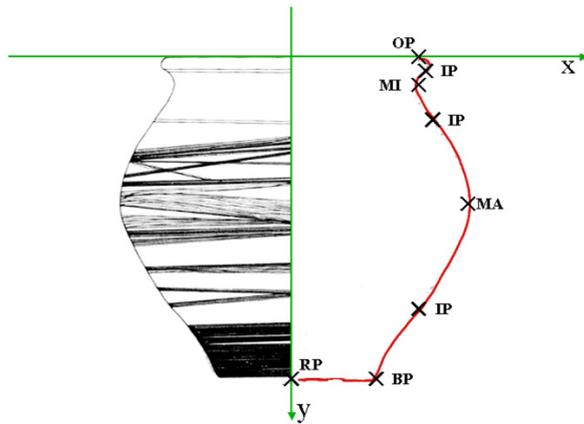


Fig. 8. S-shaped vessel: profile segmentation scheme.

- MI, local minimum: point of vertical tangency; point where the x -value is smaller than in the surrounding area of the curve.
- MA, local maximum: point of vertical tangency; point where the x -value is bigger than in the surrounding area of the curve; the y -value refers to the height of the object (e.g. $MA(y)$).
- CP, corner point: point where the curve changes its direction substantially.
- BP, base point: outermost point, where the profile line touches the base plane.
- RP, point of the axis of rotation: point where the profile line touches the axis of rotation.
- EP, end point: point where the profile line touches the axis of rotation; applied to incomplete profiles.

The profile determined has to be converted into a parameterized curve (Hu and Ma, 1995) and the curvature has to be computed (Bennett and MacDonald, 1975). Local changes in curvature (Rosenfeld and Nakamura, 1997) are the basis for rules required for segmenting the profile. Based on the B-spline methods (Hall and Laffin, 1984) the profile is thus converted into one or more mathematical curves.

In order to apply interpolation and approximation methods the profile is subdivided into sub-intervals by using corner points. The most appropriate interpolation and approximation methods are computed and selected for each of the intervals of the curve, the method with a smaller error (in case of ambiguity, the interpolation method is preferred) is selected for the interval. The approximation error of the representation over the whole curve is computed via the sum of squares of the differences of the input value and the spline value. The situation is complicated by the fact that ceramic vessels, produced by hand, do not have mathematically perfect surfaces which affects the application of the above mentioned methods. Consequently, the precision of the representation of the vessels is reduced (Rice, 1987). The technique applied is thoroughly described in (Kampel and Sablatnig, 2002).

4. Generation of primitives

The attributes of a successful classification have been summarized by Orton et al. (1993):

- objects belonging to the same type should be similar (internal cohesion),
- objects belonging to different types should be dissimilar (external isolation),
- the types should be defined with sufficient precision to allow others to duplicate the classification,
- it should be possible to decide to which type a new object belongs.

In order to achieve these aims our classification scheme of the vessel form is based on two aspects (Orton et al., 1993):

- absolute measurements and ratios,
- segmentation of the profile line.

The first step is the measurement of the following parameters: rim diameter, bottom diameter, height, x - and y -values of all segmentation points. With these measurements a variety of ratios can be calculated. A specific choice of these ratios is in each case characteristic for one vessel type; for example the ratio rim diameter to the height of the fragment.

The characteristic points together with the following measurements are used to define basic vessel forms and types (Andraschko et al., 1990):

- Rim-diameter, rdm : The diameter of the orifice plane.
- Wall-diameter, wdm : The maximum diameter of the object orthogonal to its rotational axis.
- Bottom-diameter, bdm : Diameter at the bottom of the object.
- Height, h : The overall height of the object.
- The characteristic ratio, r_{chr} : The ratio between height h and rim-diameter rdm : $h:rdm$.

Together with archaeologists (Adler et al., 2001) three levels of hierarchical classification rules based on the work of Schreg (1978) and Andraschko et al. (1990) have been worked out. They consist of three consecutive levels *ware*, *basic form* and *basic type*, see Table 1. These rules were applied to the late Roman burnished ware of Carnuntum (Gruenewald, 1986). The first classification level defines the fabric or pottery ware.

Table 1
Three levels of classification

	Ware	Late Roman burnished ware
I	Basic form	Beaker, plate, bowl, pot, jug
II	Basic type	Beaker1, beaker2, pot1, pot2, plate1–2, plate11–2, plate13–7, jug1, jug2–3, jug4

Table 2
Classification level II: specific vessel forms

Basic vessel form	r_{chr} ($\pm 15\%$)	rdm (cm)	wdm (cm)
Plate	1:8	16–34	–
Bowl	1:2–1:4	10–16 12–30	–
Beaker	1:1	8–11	5–14
Jug	4:1–2:1	6–14	–
Pot	1:1–3:1	8–12 12–16	15–25 18–21

Classification level II defines *basic forms* (see Table 2). The grouping follows functional aspects based on characteristic ratios and diameter. A variation of $\pm 15\%$ is taken

into account. For example, a plate is defined by $r_{chr} = 1:8$ and rdm ranging from 16 cm to 34 cm.

The forms are subdivided into *basic vessel types* (see Table 3), which are defined in level III. The grouping follows the characteristic properties of the profile section and the position of the characteristic points. Table 3 shows the rules and images of all vessel types which are taken into account ($a \gg b$ means at least 20% difference between a and b). These rules were found empirically and are only valid for the specific ware. References of all forms to Gruenewald (1986) and representative images of the specific form are given. For example, a plate is further specified by not having an inflection point IP and no curvature point CP.

Table 3
Classification level III: basic vessel types

Basic vessel types	Characteristic points	Characteristic properties of the profile section	Reference (Gruenewald, 1986)	Image
Beaker1	1 IP, no CP	$rdm \gg bdm$	78/1–4	
Beaker2	>1 IP	$rdm \approx wdm$	78/6	
Plate1–2	1 IP, no CP	$f(x+1) > f(x)$	71/9; 75/1–4	
Bowl1–2	1 IP, no CP	$f(x+1) \geq f(x), MA(y) < \frac{h}{10}$	70/1–6; 71/2	
Bowl3–7	CP, >1 IP	$f(x+1) \gg f(x)$	72/5–8; 73/1–3; 74/4, 6–8	
Pot1	>1 IP	$rdm \geq bdm, rdm \ll wdm, MA(y) > \frac{h}{5}$	79/2; 81/2	
Pot2	1 IP, no CP	$rdm \ll wdm, MA(y) \cong \frac{h}{2}$	79/1, 3	
Jug1	>1 IP	$rdm < wdm, wdm \gg bdm$	–	
Jug2–3	CP	$rdm \approx bdm$	84/10	
Jug4	>1 IP or CP	$rdm < 12 \text{ cm}$	84/1–3, 8; 85/2–7, 9–11	

5. Results

First we show results of finding characteristic points using the multi-spline method. Next we present experiments on the segmentation of the characteristic points, which is followed by examples of automatic classification. The pottery dataset used for the experiments was already classified by archaeologists and published by Gruenewald (1986). Besides the manual drawings of the fragments, the fragments themselves were available. In order to evaluate the classification results achieved, we randomly selected eight fragments from the pottery dataset, classified them, and compared the results to manual classification results of the same fragments.

Following the manual approach of the archaeologists, we take the outer profile line as basis for the segmentation: Fig. 9 shows two examples of automatically segmented pots with the characteristic points detected shown in (b), and the appropriate manual segmentation in (a).

Applying level II of the classification scheme to the curves computed gives a first indication of the group to which the object belongs. Table 4 summarizes results from eight fragments. The most important measurement is the diameter of the rim rdm , because its estimation does not depend on whether the object is a fragment or a vessel. On the contrary the diameter of the bottom bdm represents only a reliable indicator for vessels, because the bdm of a fragment is not the same as the bdm of the whole vessel.

For example, the rdm of fragment 70/1 is 15.6 cm, which allows the forms *plate*, *bowl* and *pot*. Its characteristic ratio $r_{chr} = 1: < 1.5$ excludes *plates*, leading to *bowl* and *pot* as indication. Ambiguities are resolved within classification level III.

Table 5 shows results of the estimation of characteristic points. Except for IP, where the number of IPs found is shown, the coordinates of the characteristic points are given. One of the most reliable indicators is the number of IPs, because it identifies S-shaped forms by simply counting points without requiring of their position. The y -position of the MA characterizes forms reliably, because its approximate position is sufficient.

To continue the previous example with fragment 70/1, the forms *bowl* and *pot* are further investigated: the position of MA indicates that it is not a pot, since it lies within

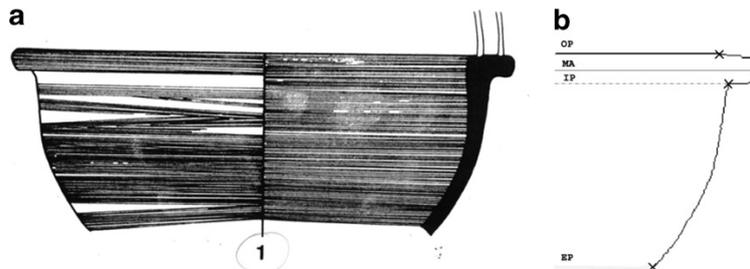


Fig. 9. Classified pottery 70/1: (a) manual drawings, (b) detected characteristic points.

Table 4

Classification level II: characteristic forms

ID (Gruenewald, 1986)	rdm	wdm	bdm	h	r_{chr}	Indication
70/1	15.6	19.2	9.2	10.2	1:1.5	Bowl, pot
70/4	19.6	21.6	9.6	6.4	1:3.0	Bowl
72/6	14.4	17.0	5.2	7.4	1:1.9	Bowl, pot
72/8	22.2	24.2	7.2	10.0	1:1.2	Bowl
75/3	29.6	32.4	16.2	5.6	1:5.3	Plate
78/2	6.4	7.2	4.6	9.2	1.4:1	Jug, pot, beaker
79/2	12	14.0	5.2	10.6	1:1.1	Pot, beaker
81/1	12.4	15.4	4.8	11.3	1:1.1	Beaker, pot

Table 5

Classification level III: characteristic types

ID	OP	MA	MI	# IP	CP	BP	RP	EP
70/1	[7.8,0.0]	[9.6,0.8]	EP	1	–	–	–	[4.6,10.2]
70/4	[9.8,0.0]	[10.8,0.8]	BP	1	–	[4.8,6.4]	–	–
72/6	[7.2,0.0]	[8.5,1.0]	EP	3	–	–	–	[2.6,7.4]
72/8	[11.1,0.0]	[12.1,0.9]	EP	3	–	–	–	[3.6,10.0]
75/3	[14.8,0.0]	[16.2,0.8]	BP	1	–	[8.1,5.6]	–	–
78/2	[3.2,0.0]	[3.6,0.5]	EP	2	–	–	–	[2.3,9.2]
79/2	[6.0,0.0]	[7.0,1.5]	EP	1	–	–	–	[2.6,10.6]
81/1	[6.2,0.0]	[7.7,1.8]	EP	1	–	–	–	[2.4,11.3]

the first 20% of the height h . Having 1 IP only gives priority to Bowl1–2, which actually is a correct classification (Gruenewald, 1986).

Table 6

Final classification

ID (Gruenewald, 1986)	Level II	Classification rules applied	Indication
70/1	Bowl, pot	1 IP, $MA(y) < \frac{h}{5} (\cong 0.8 < 2.0)$	Bowl1–2
70/4	Bowl	1 IP, $MA(y) < \frac{h}{5} (\cong 0.8 < 1.3)$	Bowl1–2
72/6	Bowl, pot	3 IP, $MA(y) < \frac{h}{5} (\cong 1.0 < 1.5)$	Bowl3/7
72/8	Bowl	3 IP, $MA(y) < \frac{h}{5} (\cong 0.9 < 2.0)$	Bowl3/7
75/3	Plate	1 IP	Plate1–2
78/2	Jug, pot	2 IP, $rdm > bdm; MA(y) < \frac{h}{5} (\cong 0.5 < 1.8)$	Beaker1
79/2	Pot, beaker	1 IP, $MA(y) < \frac{h}{5} (\cong 1.5 < 2.1)$	Pot1, pot2
81/1	Pot, beaker	1 IP, $MA(y) < \frac{h}{5} (\cong 1.8 < 2.3)$	Pot1, pot2

Table 6 summarizes the results for the eight randomly selected fragments. The results achieved indicate the same basic vessel forms as published by Gruenewald (1986), except for 79/2 and 81/1. The reason is that distinction between forms pot1 and pot2 is not possible for fragments, where the bottom is missing, because it depends only on the relative position of MA.

6. Conclusion

In this paper, we presented the classification of the profile section based on the manual approach of the archaeologists. In order to define shape characteristics, a classification scheme was defined. The profile was segmented based on local changes in curvature, therefore B-Splines have been applied to the profiles. Three levels of classification based on characteristic measurements and the segmentation of the profile allowed the grouping of the fragment into specific vessel forms and specific vessel types. Results were shown for each intermediate step. The classification scheme presented depends on the excavation site. In order to apply our algorithm to other fragment-fabrics we have to redefine the classification scheme based on the characteristics of the specific fabric. Currently we are recording fragments at the excavation site Tel Dor in Israel. It is planned to apply our classification system on the resulting 3D-models (150–200 fragments).

References

- Adams, W.Y., Adams, E.W., 1991. *Archaeological Typology and Practical Reality. A Dialectical Approach to Artifact Classification and Sorting*. Cambridge.
- Adler, K., Kampel, M., Kastler, R., Penz, M., Sablatnig, R., Hlavackova-Schindler, K., 2001. Computer aided classification of ceramics – achievements and problems. In: Börner, W., Dollhofer, L. (Eds.), *Proc. 6th Internat. Workshop on Archaeology and Computers*, Vienna, Austria, pp. 3–12.
- Andraschko, F.M., Krekeler, A., Schliep-Andraschko, N., 1990. Terminologie, Klassifikation und computergestützte Bearbeitung der Keramik des 1. Jahrtausends v. Chr. vom Westkom auf Elephantine, Oberägypten. In: Teegen, W.R., Andraschko, F.M. (Eds.), *Gedenkschrift für Jürgen Driehaus*, pp. 327–338.
- Bennett, J.R., MacDonald, J.S., 1975. On the measurement of curvature in a quantized environment. *IEEE Trans. Comput.* 24, 803–820.
- Besl, P.J., 1988. Active, optical range imaging sensors. *Machine Vision Appl.* 1 (2), 127–152.
- Binford, L.R., 1965. Archaeological systematics and the study of cultural process. *Amer. Antiquity* 31, 203–210.
- Degeest, R., 2000. The common wares of sagalassos. In: Walkens, M. (Ed.), *Studies in Eastern Mediterranean Archaeology*, III.
- DePiero, F., Trivedi, M., 1996. 3D computer vision using structured light: Design, calibration, and implementation issues. *Adv. Comput.*
- Gilboa, A., Karasik, A., Sharon, I., Smilansky, U., 2004. Towards computerized typology and classification of ceramics. *J. Archaeol. Sci.* 31, 681–694.
- Gruenewald, M., 1986. *Ausgrabungen im Legionslager von Carnuntum (Grabungen 1969–1977). Der römische Limes in Österreich* 34, 10–11.
- Hall, N.S., Laffin, S., 1984. A computer aided design technique for pottery profiles. In: Laffin, S. (Ed.), *Computer Applications in Archaeology*. Computer Center, University of Birmingham, Birmingham, pp. 178–188.
- Hu, Z., Ma, S.D., 1995. The three conditions of a good line parameterization. *Pattern Recognition Lett.* 16, 385–388.
- Kampel, Martin, Sablatnig, Robert., 2003. Profile based pottery reconstruction. In: Martin, D. (Ed.), *Proc. IEEE/CVPR Workshop on Applications of Computer Vision in Archaeology*, Madison, Wisconsin, USA, pp. CD-ROM.
- Kampel, M., Sablatnig, R., 2002. Automated Segmentation of archaeological profiles for classification. In: Kasturi, R., Laurendeau, D., Suen, C. (Eds.), *Proc. 16th Internat. Conf. on Pattern Recognition*, Quebec City, vol. 1. IEEE Computer Society, pp. 57–60.
- Orton, C., Tyers, P., Vince, A., 1993. *Pottery in Archaeology*. Cambridge University Press, Cambridge.
- Poblome, J., 1999. Sagalassos red slip ware. In: Walkens, M. (Ed.), *Studies in Eastern Mediterranean Archaeology*, II, Brepols.
- Rice, P.M., 1987. *Pottery Analysis: A Sourcebook*.
- Rosenfeld, A., Nakamura, A., 1997. Local deformations of digital curves. *Pattern Recognition Lett.* 18 (7), 613–620.
- Sablatnig, R., Kampel, M., 2002. Model-based registration of front- and backviews. *Computer Vision and Image Understanding* 87 (1), 90–103.
- Schreg, R., 1978. *Keramik aus Südwestdeutschland. Eine Hilfe zur Beschreibung, Bestimmung und Datierung archäologischer Funde vom Neolithikum bis zur Neuzeit*, Tübingen.
- Sinopoli, C.M., 1991. *Approaches to Archaeological Ceramics*, New York.

Chapter 6

Recognizing Ancient Coins based on Local Features

Martin Kampel and Maia Zaharieva. Recognizing Ancient Coins based on Local Features, In G. Bebis et al., editors, *Advances in Visual Computing*. Springer. LNCS 5358, pp. 11-22, 2008.

Recognizing Ancient Coins Based on Local Features*

Martin Kampel and Maia Zaharieva

Pattern Recognition and Image Processing Group – TU Vienna
Favoritenstr 9 - 11, 1040 Vienna, Austria
kampel@prip.tuwien.ac.at

Abstract. Numismatics deals with various historical aspects of the phenomenon money. Fundamental part of a numismatists work is the identification and classification of coins according to standard reference books. The recognition of ancient coins is a highly complex task that requires years of experience in the entire field of numismatics. To date, no optical recognition system for ancient coins has been investigated successfully. In this paper, we present an extension and combination of local image descriptors relevant for ancient coin recognition. Interest points are detected and their appearance is described by local descriptors. Coin recognition is based on the selection of similar images based on feature matching. Experiments are presented for a database containing ancient coin images demonstrating the feasibility of our approach.

1 Introduction

Numismatics is at a point where it can benefit greatly from the application of computer vision methods, and in turn provides a large number of new, challenging and interesting conceptual problems and data for computer vision. For coin recognition we distinguish between two approaches: coin identification and coin classification. A coin classification process assigns a coin to a predefined category or type, whereas a coin identification process assigns a unique identifier to a specific coin. What makes this application special and challenging for object recognition, is that all the coins are very similar.

The first coins were struck in Asia Minor in the late 7th century BC. Since then coins are a mass product [1]. In the Antiquity coins were hammer-struck from manually engraved coin dies. Coins from the same production batch will have very much the same picture and also the same quality of its relief. Depending on the series of coins in question, the only varying details can be either part of the picture or legend or there can be a difference in a prominent detail such as the face of a figure. The scientific requirement is to assign a coin its correct number according to a reference book.

* This work was partly supported by the European Union under grant FP6-SSP5-044450. However, this paper reflects only the authors' views and the European Community is not liable for any use that may be made of the information contained herein.



Fig. 1. Different coins of the same coin type



Fig. 2. Different image representations of the same coin

Ancient and modern coins bear fundamental differences that restrict the applicability of existing algorithms [14]. Due to their nature ancient coins provide a set of identifying features. The unique shape of each coin originates in the manufacturing process (hammering procedure, specific mint marks, coin breakages, die deterioration, etc.). Furthermore, the time leaves its individual mark on each coin (fractures, excessive abrasion, damage, corrosion, etc.). Eventually, identification of ancient coins turns out to be "easier" compared to classification. For example, Figure 1 shows ten different coins of the same coin type. A classification algorithm should ideally classify them all of the same class. However, they all provide complete different characteristics (see shape, die position, mint marks or level of details). At the same time, exactly those features enable the identification process.

In contrast, Figure 2 presents five pictures of one and the same coin. The pictures were taken using different acquisition setups, i.e. scan as well as fixed and free hand cameras with varying lighting conditions. The figure points out the challenges for an automated identification process as well as the importance of quality images for the process itself. Different lighting conditions can hide or show details on the coin that are significant for a successful identification process (e.g. compare the first and the third image in Figure 2).

The remainder of this paper is organised as follows: In Section 2 related work on recognizing coins is presented. Section 3 gives an overview of local features with respect to our needs. The coin recognition workflow is described in Section 4. The experiments performed and their results are presented in Section 5. We conclude the paper in Section 6 with discussion on the results achieved and an outlook for further research.

2 Related Work

Research on pattern recognition algorithms for the identification of coins started in 1991 when Fukumi et al. [2] published their work on rotation-invariant visual coin recognition using a neural networks approach. Also [3] is devoted to neural network design, but investigates the possibilities of simulated annealing and genetic algorithms. In 1996 Davidson [4] developed an instance-based learning algorithm based on an algorithm using of decision trees [5]. An industrial implementation of a neural networks approach is described in [6].

A more recent neural algorithm was published in [7]. This approach employs the output of a filter bank of Gabor filters fed into a back propagation network. The algorithm uses correlation in the polarspace and in combination with a neural networks. Khashman et al. implemented a neural coin recognition system for use in slot machines [8].

Huber et al. present in [9] a multistage classifier based on *eigenspaces* that is able to discriminate between hundreds of coin classes. The *Dagobert* coin recognition system presented by Nölle et al. [10] aims at the fast classification of a large number of modern coins from more than 30 different currencies. In their system coin classification is accomplished by correlating the edge image of the coin with a preselected subset of master coins and finding the master coin with lowest distance.

In [11] Maaten et al. present a coin classification system based on *edge-based statistical features*. It was developed for the MUSCLE CIS Coin Competition 2006 [12] focusing on reliability and speed. The coin classification method proposed by Reisert et al. [13] is based on *gradient information*. Similar to the work of Nölle et. al [10] coins are classified by registering and comparing the coin with a preselected subset of all reference coins.

Current research approaches for coin recognition algorithms possess mainly two limitations. On the one hand, the input digital image is well defined – there is always only one coin pictured and the image is taken under very controlled conditions (such as background, illumination, etc.). On the other hand, the algorithms focus mainly on the recognition of modern coins. Those assumptions facilitate the classification and identification process substantially. In the case of controlled conditions and the well known circular shape of modern coins, the process of coin detection and segmentation becomes an easier task. The almost arbitrary shape of an ancient coin narrows the amount of appropriate segmentation algorithms. Tests performed on image collections both of medieval and modern coins show that algorithms performing good on modern coins do not necessarily meet the requirements for classification of medieval ones [14]. The features that most influence the quality of recognition process are yet unexplored.

3 Local Image Features

Local features describe image regions around given interest points. Their application in the computer vision is manifold ranging from object and texture recognition [15] to robot localization [16], symmetry detection [17] and wide

baseline stereo matching [18]. Local features are already successfully used for object classification. Crucial influence on local feature based object recognition bear both the detection of interest points and their representation. Hence, in the following we give a short overview over top performing interest point detectors and local feature descriptors and discuss their applicability with respect to the identification of ancient coins.

3.1 Interest Point Detectors

In the literature exist a broad number of interest point detectors with varying level on invariance against rotation, scale or affine changes. Comparative studies on interest points and their performance evaluation can be found in [19,20].

The *Harris* corner detector [21] is based on local auto-correlation matrix of the image function. The squared first derivatives are averaged over a 41×41 Gaussian weighted window around an image point. If the auto-correlation matrix has two significant eigenvalues, an interest point is detected. However, the detected points are not invariant to scale and affine changes. To achieve scale invariance Mikolajczyk et al. [22] extend the Harris detector by selecting corners at location where a Laplacian attains an extrema in scale space (*Harris-Laplace*). The *Harris-Affine* detector [22,19] additionally uses second moment matrix to achieve affine invariance. Detected points are stable under varying lighting conditions since significant signal change in orthogonal directions is captured.

Hessian-Laplace localizes points at local maxima of the Hessian determinant in scale-space maxima of the Laplacian-of-Gaussian [15,19]. Detected keypoints are invariant to scale and rotation transformations. Similar to Harris-Affine, the *Hessian-Affine* detector provides in a next step affine invariance based on second moment matrix [19]. In contrary to the Harris-based detectors, Hessian interest points indicate the presence of blob like structures. Bay et al. [23] introduced recently a further detector based on the Hessian matrix – the *Fast-Hessian* detector. It approximates Gaussian second order derivative with box filter. To further reduce the computational time, image convolutions use integral images.

Tuytelaars et al. present in [18] further two methods to extract affine invariant regions. The *Geometry-based region* detector starts from Harris corners and uses the nearby edges identified by the Canny edge operator [24] to build a parallelogram. Keypoints are detected if the parallelogram goes through an extremum of intensity-based functions. The second method proposed – *Intensity-based region* detector – relies solely on the analysis of image intensity. It localizes interest points based on intensity function along rays originating from local extrema in intensity.

The *Maximally Stable Extremal Regions* (MSER) proposed by Matas et al. [25] are a watershed based algorithm. It detects intensity regions below and above a certain threshold and select those which remain stable over a set of thresholds.

The *Difference-of-Gaussian* (DoG) detector was introduced by Lowe as keypoint localization method for the Scale Invariant Feature Transform (SIFT) approach [26,15]. Interest points are identified at peaks (local maxima and minima) of Gaussian function applied in scale space. All keypoints with low contrast or keypoints that are localized at edges are eliminated using a Laplacian function.

Table 1. Average interest points detected

Detector	Interest points
Difference-of-Gaussian (DoG) [26]	968
Harris-Laplace [15]	204
Harris-Affine [19]	198
Hessian-Laplace [19]	1076
Hessian-Affine [19]	778
Fast-Hessian [23]	198
Geometry-based region (GBR) [18]	61
IBR [18]	184
Maximally Stable Extremal Regions (MSER) [25]	134

Common critic to edge-based methods is that it is more sensitive to noise and changes in neighboring texture. Interest point detectors which are less sensitive to changes in texture perform well in a classification scenario since they recognize and capture those features that are common for all instances in a given class. On the opposite, identification relies on those features that are unique for a given object. Due to their nature and manufacturing process, ancient coins are unique. Coins produced by the same die show the same picture. However, since they are hand-hammered, shape, texture and relief can vary to a large degree. In this particular scenario, texture-sensitive interest point detectors are expected to perform better. Table 1 shows average interest points extracted per detector for the dataset explained in Section 5.

As we will show in Section 5, the methods which detect most interest points do not necessarily perform the best. First, we are faced with the problem of overfitting (i.e. each coin is similar to all the other coins to some degree). Second, essential role play the information captured per interest point. Thus, in the next subsection we give a short overview of the local feature descriptors we used for the experiments.

3.2 Local Feature Descriptors

Given a set of interest points, the next step is to choose the most appropriate descriptor to capture the characteristics of a provided region. Different descriptors emphasize different image properties such as intensity, edges or texture. Please refer to [27] for a thorough survey on the performance of local feature descriptors. We focus our study on four descriptors which show outstanding performance with respect to changes in illumination, scale, rotation and blur.

(1) Lowe [15] introduced the *Scale Invariant Feature Transform (SIFT)* descriptor which is based on gradient distribution in salient regions – at each feature location, an orientation is selected 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.

(2) Mikolajczyk and Schmid [27] propose an extension of the SIFT descriptor – *Gradient Location and Orientation Histogram (GLOH)* – designed to increase the robustness and distinctiveness of the SIFT descriptor. Instead of dividing the path around the interest points into a 4×4 grid, the authors divide it into radial and angular grid. A log-polar location grid with 3 bins in radial and 8 bins in angular directions is used. The gradient orientations are quantized into 16 bins which gives a 272 bin histogram further reduced in size using PCA to 128 feature vector dimension.

(3) Belongie et al. [28] introduce *Shape Context* as feature descriptor for shape matching and object recognition. The authors represent the shape of an object by a discrete set of points sampled from its internal or external boundaries. As starting points, edge pixels as found by an edge detector. Following, for each point the relative location of the remaining points is accumulated in a coarse log-polar histogram.

(4) *Speeded Up Robust Features (SURF)* [23] 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.

In [27] complementary evaluation on the performance of local descriptors with respect to rotation, scale, illumination, and viewpoint change, image blur and JPEG compression, is presented. In most of the tests SIFT and GLOH clearly outperformed the remaining descriptors: shape context, steerable filters, PCA-SIFT, differential invariants, spin images, complex filters, and moment invariants. In [29] Stark and Schiele report that the combination of Hessian-Laplace detector with SIFT and GLOH descriptor outperforms local features such as Geometric Blur, k-Adjacent Segments and Shape Context in a object categorization scenario. For their evaluation the authors used three different datasets containing quite distinguishable objects such as cup, fork, hammer, knife, etc. By contrast, our two coin data sets possess very different characteristics in comparison with existing evaluation and application scenarios. Both data sets contain similar objects and both are targeted to evaluate identification performance.

4 Recognition Workflow

We define the workflow for the identification of ancient coins by five well-defined stages as shown in Figure 3. In the *preprocessing* step (1) coins contained in the image are detected and segmented. Essential influence on the process carries the image diversity, e.g. single or multiple objects pictured, varying lighting conditions, shadows, diverse background textures, etc. In the scenario of ancient coins identification the almost arbitrary shape of coin additionally impede the task of coin(s) detection and segmentation. Since our test database consists solely of images of single coin on an unitary background no preprocessing is required. Eventually, the applied local feature detectors locate interest points on the background (e.g. due to intensity change). However, their amount is minimal and has no influence on the identification process.

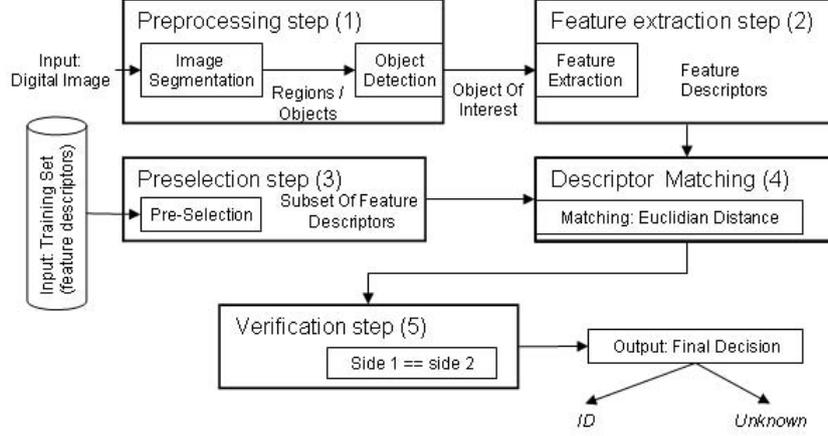


Fig. 3. The five stages of coins identification workflow

The goal of the *feature extraction* step (2) is twofold. First, local features algorithms are applied to extract local image descriptors for coins identification. Second, features that can be used to reduce the number of required feature comparisons by reducing the coins database can be extracted. Provided uncontrolled acquisition process, simple features such as area or perimeter of a coin are not eligible since the scaling factor is unknown. Other features such as shape descriptors can be used as basis for step (3) - *preselection* step [30].

Step (4) *descriptor matching* is performed by identifying the first two nearest neighbors in terms of Euclidean distances. A descriptor D_1 is accepted only if the distance ratio of the nearest (1.NN) to the second nearest (2.NN) neighbors is less than or equal to 0.5:

$$2d(D_1, D_{1.NN}) \leq (D_1, D_{2.NN}). \quad (1)$$

In [15] Lowe suggests a distance ratio of 0.8. However, our experiments showed that for the case of lower inter-class differences (as all classes are coins), a lower distance ratio tends to keep more distinctive descriptors while eliminating a great part of the false matches. The value of 0.5 was determined experimentally and used throughout the tests. Furthermore, we apply a restriction rule to fasten the quality of the matches. Since each image in the database picture is a single ancient coin, a given keypoint can only be matched to a single point in a different feature set. Thus, all multiple matches are removed as they are considered to be unstable for the identification process.

Finally, an additional *verification* step (5) can assure the final decision. Provided images of both obverse and reverse side of a coin, each side is first identified separately. If both sides vote for the same coin identification, the coin is identified adequately. Otherwise, it is classified as unknown.

5 Experiments

For our experiments we used a dataset of images acquired at the Fitzwilliam Museum in Cambridge, UK. We used varying technical setups – scan as well as fixed and free hand cameras, and varying lighting conditions. The dataset consists of 350 images of three different coin types (10 to 16 coins à coin type, 3 to 5 pictures à coin side). Ground truth is encoded in the file names. For testing the recognition one image was selected as test images. Presented evaluations as [27,29] on the performance of local descriptors use different datasets containing quite distinguishable objects such as cup, fork, hammer, knife, etc. By contrast, our coin data set possesses very different characteristics in comparison with existing evaluation and application scenarios. The data set contains similar objects and is targeted to evaluate coin recognition performance.

In a first experiment we compare the performance of three descriptors on coin identification.

Figure 4 shows corresponding interest points detected by the different approaches. Despite the lower image quality of the input image, the rotation and scale change of the coin, the SIFT approach matches correctly against image of the same coin acquired by the scan device (see Figure 4(a)).

The Fast Approximated SIFT approach – Figure 4(b) – tends to detect keypoints mostly on the background of the image. The algorithm detects far more points than SIFT, e.g. for the example input image 8999 keypoints (by contrast keypoints detected by SIFT for the same image: 721). However, they lack of stability and distinctiveness. Eventually, each detected interest point is similar (i.e. being matched) to a large number of keypoints in the second image. The elimination of multiple matched points reduces the number of final matches by approximately 90%.

Performing manual pairwise comparison of the resulting matches, PCA-SIFT (see Figure 4(c)) seems to achieve almost the same amount on descriptors as SIFT for less computational time. However, the stability of the PCA-SIFT features is considerably lower since approximately 40% of the correct classified

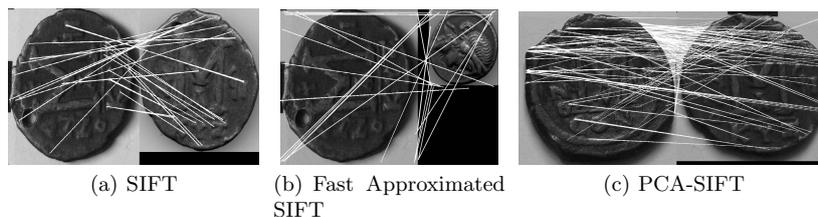


Fig. 4. Example matches for a given ancient coin acquired using a free hand camera (Input image on the left side and corresponding match on the right hand). Using the SIFT approach (a), the test coins was successfully matched against an image of the same coin acquired using scan device. The Fast Approximated SIFT fail to recognize the image (b). PCA-SIFT (c) matched against different coin of the same coin type.

Table 2. Evaluation results on the recognition performance of the local image feature descriptors using the small database of ancient coins. CR shows the rate of correctly *classified* coins, and IR those of correctly *identified* ancient coins.

Interest Point Detectors	(1) SIFT		(2) GLOH		(3) Shape		(4) SURF	
	CR	IR	CR	IR	CR	IR	CR	IR
DoG	90.57%	84.57%	60.00%	40.00%	61.14%	29.14%	82.57%	28.57%
Harris-Laplace	68.39%	50.86%	71.84%	53.45%	79.71%	61.45%	71.30%	28.12%
Harris-Affine	76.15%	55.46%	73.56%	54.31%	73.04%	53.04%	71.88%	27.83%
Hessian-Laplace	65.90%	47.28%	65.90%	47.28%	92.57%	82.00%	84.29%	32.29%
Hessian-Affine	71.63%	50.72%	68.48%	49.28%	88.00%	80.00%	79.43%	29.71%
Fast-Hessian	85.43%	79.43%	85.43%	78.29%	84.86%	72.29%	90.86%	78.29%
GBR	51.47%	27.36%	48.53%	24.76%	52.44%	29.64%	56.03%	15.31%
IBR	80.29%	60.57%	75.71%	50.57%	80.29%	55.14%	77.43%	25.14%
MSER	80.86%	68.29%	77.71%	64.29%	74.86%	58.00%	74.00%	28.29%

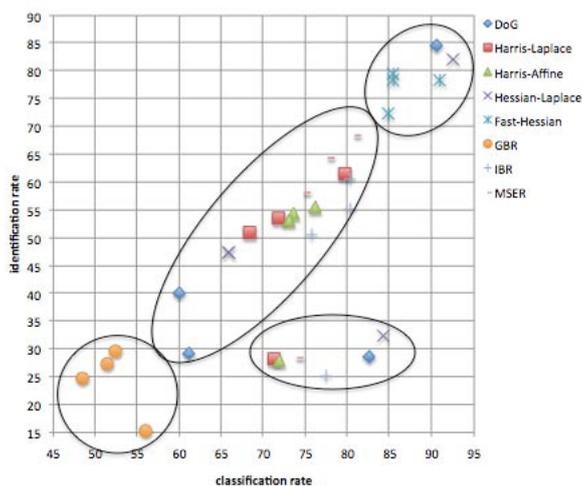


Fig. 5. Performance distribution of the interest point detectors

images are due matching of the obverse with the reverse side of a coin. The PCA reduction of feature vector size seems to lead to loss of valuable information for the identification process. In terms of identification rate, SIFT clearly outperforms both modifications by more than 10%.

The second experiment aims at evaluation of the performance of the presented interest point detectors and local descriptors with respect to recognition. We compare both classification (CR) and identification rate (IR) and show that a good classification rate is no guarantee for the distinctiveness and stability of the respective detectors or descriptors. Table 2 summarizes the results on the

coin data set. The best classification rate of 92.57% was achieved with Shape Context combined with Hessian Laplace detector. The best identification rate of 84.57% was achieved with SIFT combined with DoG. The main reason for the significant difference between classification and identification rate is the nature of local descriptors. Local descriptors simply describe the close surroundings of given interest point. Dependent on the size of this box, matching, i.e. similar enough, descriptor can be found on multiple coins or even on the same coin or different sides of the same coin.

Figure 5 visualizes the performance distribution with respect to the interest point detectors. One can clearly identify four groups. The first one, low identification and low classification rate, is dominated by the GBR detector. Independent of the applied local feature descriptor the achieved performance is too low with a rate close or far below 50%. The second group, high classification and low identification rate, is defined by the use of the SURF descriptor. Independently of the applied interest point detector, the SURF descriptor shows high stability with respect to classification. The last conspicuous group, high classification and high identification rate, is dominated by the Fast Hessian detector.

6 Conclusion

In this paper, we described a strategy for the recognition of ancient coins based on local image features. The achieved recognition rates indicate the feasibility of the approach. SIFT features show outstanding performance in existing evaluations. However, the main drawback and critical point is their computational time. Benefits of the proposed system are in the field of coin recognition. Based on the promising results we plan to extend the evaluation on a recently recorded coin collection of 2400 images of 240 different coins. Future research will include methods in the field of optical character and symbol recognition. Furthermore, we will extend our work towards die and mint sign identification based on spatially constrained local features.

Acknowledgment. The authors want to thank Dr. Mark Blackburn and his team at the Coin Department, Fitzwilliam Museum, Cambridge, UK, for sharing their experience and providing the database of images of ancient coins.

References

1. Duncan-Jones, R.: Money and Government in the Roman Empire. Cambridge (1994)
2. Fukumi, M., Omatu, S., Takeda, F., Kosaka, T.: Rotation-invariant neural pattern recognition system with application to coin recognition. *IEEE Transactions on Neural Networks* 3, 272–279 (1992)
3. Mitsukura, Y., Fukumi, M., Akamatsu, N.: Design and evaluation of neural networks for coin recognition by using ga and sa. In: *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, IJCNN 2000*, vol. 5, pp. 178–183 (2000)

4. Davidsson, P.: Coin classification using a novel technique for learning characteristic decision trees by controlling the degree of generalization. In: Proc. of 9th Int. Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems (IEA/AIE 1996), pp. 403–412 (1996)
5. Aha, D., Kibler, D., Albert, M.: Instance-based learning algorithms. *Machine Learning* 6, 37–66 (1991)
6. Moreno, J.M., Madrenas, J., Cabestany, J., Laúna, J.R.: Using classical and evolutive neural models in industrial applications: A case study for an automatic coin classifier. In: *Biological and Artificial Computation: From Neuroscience to Neurotechnology*, pp. 922–931 (1997)
7. Bremananth, R., Balaji, B., Sankari, M., Chitra, A.: A new approach to coin recognition using neural pattern analysis. In: *Proceedings of IEEE INDICON 2005*, pp. 366–370 (2005)
8. Khashman, A., Sekeroglu, B., Dimililer, K.: Rotated Coin Recognition Using Neural Networks. *Advances in Soft Computing*, vol. 41, pp. 290–297. Springer, Berlin (2007)
9. Huber, R., Ramoser, H., Mayer, K., Penz, H., Rubik, M.: Classification of coins using an eigenspace approach. *Pattern Recognition Letters* 26, 61–75 (2005)
10. Nölle, M., Penz, H., Rubik, M., Mayer, K.J., Holländer, I., Granec, R.: Dagobert – a new coin recognition and sorting system. In: *Proc. of the 7th International Conference on Digital Image Computing - Techniques and Applications (DICTA 2003)*, Macquarie University, Sydney, Australia, pp. 329–338. CSIRO Publishing (2003)
11. van der Maaten, L.J., Poon, P.: Coin-o-matic: A fast system for reliable coin classification. In: *Proc. of the Muscle CIS Coin Competition Workshop*, Berlin, Germany, pp. 7–18 (2006)
12. Nölle, M., Rubik, M., Hanbury, A.: Results of the muscle cis coin competition 2006. In: *Proceedings of the Muscle CIS Coin Competition Workshop*, Berlin, Germany, pp. 1–5 (2006)
13. Reiser, M., Ronneberger, O., Burkhardt, H.: An efficient gradient based registration technique for coin recognition. In: *Proc. of the Muscle CIS Coin Competition Workshop*, Berlin, Germany, pp. 19–31 (2006)
14. Zaharieva, M., Kampel, M., Zambanini, S.: Image based recognition of ancient coins. In: Kropatsch, W.G., Kampel, M., Hanbury, A. (eds.) *CAIP 2007*. LNCS, vol. 4673, pp. 547–554. Springer, Heidelberg (2007)
15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 91–110 (2004)
16. Murillo, A.C., Guerrero, J.J., Sagüés, C.: Surf features for efficient robot localization with omnidirectional images. In: *IEEE International Conference on Robotics and Automation*, pp. 3901–3907 (2007)
17. Loy, G., Eklundh, J.O.: Detecting symmetry and symmetric constellations of features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3952, pp. 508–521. Springer, Heidelberg (2006)
18. Tuytelaars, T., Van Gool, L.: Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision* 59, 61–85 (2004)
19. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. *International Journal of Computer Vision* 65, 43–72 (2005)
20. Schmid, C., Mohr, R., Baukhage, C.: Evaluation of interest point detectors. *International Journal of Computer Vision* 2, 151–172 (2000)

21. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Conference, pp. 147–152 (1988)
22. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. *International Journal of Computer Vision* 60, 63–86 (2004)
23. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
24. Canny, J.: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8, 679–698 (1986)
25. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: *Proceedings of the British Machine Vision Conference, London*, vol. 1, pp. 384–393 (2002)
26. Lowe, D.G.: Object recognition from local scale-invariant features. In: *International Conference on Computer Vision (ICCV 1999)*, Washington, DC, USA, vol. 2, pp. 1150–1157. IEEE Computer Society, Los Alamitos (1999)
27. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1615–1630 (2005)
28. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 509–522 (2002)
29. Stark, M., Schiele, B.: How good are local features for classes of geometric objects. In: *11th International Conference on Computer Vision (ICCV 2007)*, Rio de Janeiro, Brazil (2007)
30. Zaharieva, M., Huber-Mörk, R., Nölle, M., Kampel, M.: On ancient coin classification. In: Arnold, D., Chalmers, A., Niccolucci, F. (eds.) *8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2007)*, Eurographics, pp. 55–62 (2007)

Chapter 7

Color Classification of Archaeological Fragments

Martin Kampel and Robert Sablatnig, Color Classification of Archaeological Fragments, In A. Sanfeliu, J. Villanueva, M. Vanrell, R. Alquezar, A. K. Jain, and J. Kittler, editors, *Proc. of 15th International Conference on Pattern Recognition*, IEEE Computer Society, Vol. 4, pp. 771-774, 2000.

Color Classification of Archaeological Fragments *

Martin Kampel and Robert Sablatnig
Vienna University of Technology
Institute for Computer Aided Automation
Pattern Recognition and Image Processing Group
Favoritenstr.9/183-2, A-1040 Vienna, Austria
{kempel,sab}@prip.tuwien.ac.at

Abstract

We are developing an automated classification and reconstruction system for archaeological fragments. The goal is to relate different fragments belonging to the same vessel based on shape, material and color, thus the color information is important in the pre-classification process. In this work a color specification technique is proposed, which exploits the fact that the spectral reflectance of materials like archaeological fragments vary slowly in the visible. We explain how the acquisition system is calibrated in order to get accurate colorimetric information with respect to archaeological requirements. Experimental results are presented for archaeological objects and for a set of test color patches.

1 Introduction

Ceramics are one of the most widespread archaeological finds and are a short-lived material. This property helps researchers to document changes of style and ornaments. Especially ceramic vessels, where shape and decoration are exposed to constantly changing fashion, not only allow a basis for dating the archaeological strata, but also provide evidence of local production and trade relations of a community as well as the consumer behavior of the local population. The purpose of ceramic classification is to get a systematic view of the material found [2, 6] and is used to relate a fragment to existing parts in the archive.

Archaeologists determine the specific color of a fragment by matching it to the Munsell color patches [7]. Since this process is done "manually" by different archaeologists and under varying light conditions, results differ from each

*This work was supported in part by the Austrian Science Foundation (FWF) under grant P13385-INF.

other. Archaeologists need digital color images of fragments for archivation purposes, thus the color information which is normally achieved with a color measurement instrument can be gained directly from the digital image for each pixel in the entire image.

We propose a solution to the color classification assuming that the spectral reflectance of archaeological fragments varies slowly in the visible spectrum. We present an approach for accurate colorimetric information on fragments, performed on digital images containing archaeological fragments under different illuminants. A characteristic vector analysis [9] of the reference reflectance leads to an algorithm that computes the colorimetrically accurate reflectance out of a video digitizing system.

The paper is organized as follows: In Section 2 we describe the theoretical background, in Section 3 we explain how we specify the colorimetric variables in order to calibrate the acquisition system with respect to archaeological requirements. Experimental results are described in Section 4 and we conclude with a summary and outline the future work.

2. Theory and Notation

Much of human color-vision research focuses on color constancy since it is the perceptual ability that permits us to discount spectral variation in the ambient light and assign stable colors: Maloney and Wandell [4] considered that both lighting and spectral reflectance are unknown, whereas Lee [3] simplified that problem by assuming that spectral illumination is known. Color and reflectance based object recognition was presented by [1, 8]. In order to provide a device-independent color specification we use reference colors from the MacBeth Color chart [5].

Our approach rests upon Lee's method assuming that spectral illumination is known and that the spectral re-

reflectance of our material varies slowly in the visible spectrum. This means that small changes of RGB values should lead to small changes in reflectance. Prior knowledge about the illuminant leads to chromaticity and luminance information.

Each RGB pixel in a digitized image has a value proportional to weighted integral over the visible spectrum. This integral depends on three spectral variables. These are the *spectral irradiance* $E(\lambda)$, which describes the energy per second at each wavelength λ . The proportion of light of wavelength λ reflected from an object is determined by the *surface spectral reflectance* $S(\lambda)$. We assume that there are k distinct channels in the digitizing system, we use $k = 3$ for red, green and blue. We denote the *spectral response* of the k th channel as $R_k(\lambda)$ and a pixel value for the k th color channel as p_k .

$$p_k = \int S(\lambda)E(\lambda)R_k(\lambda)d(\lambda) \quad (1)$$

Eq 1 describes the relationship between pixel values and spectral quantities. We approximate the three integrals above as summations over wavelength, using values every $10nm$ in the visible spectrum from $400nm$ to $700nm$. If the proportionality factor in the $R_k(\lambda)$ is subsumed, one can construct the following matrix equation (Eq. 2). m denotes the steps to be taken in the spectrum.

$$p = SER \quad (2)$$

$p \dots 1$ by 3 row vector (RGB pixel)
 $S \dots 1$ by m row vector,(surface reflectance)
 $E \dots m$ by m diagonal matrix, (spectral irradiance)
 $R \dots m$ by 3 matrix, (system spectral transfer function)

If we know elements of two of the arrays on the right side of Eq. 2 and the corresponding RGB pixel values on the left side, we can solve the unknown array. Since only an approximated knowledge of the system function R is assumed, the goal will be to:

- specify the system transfer function R more accurately by analyzing color samples with known reflectance of the MacBeth Color patches.
- use this new information to find the unknown spectral reflectance of other samples illuminated by the same light source.

The goal of the first step is to improve the transfer function R which leads to R_{new} (Eq. 3).

$$R_{new} = RR_1 \quad (3)$$

Therefore we digitize an image of the color chart, which is illuminated by the same light source that will be used

when we evaluate unknown color samples. The digitization gives a q by 3 -matrix P containing RGB values, where q denotes the number of patches of the color checker. Since we know the illumination E and the set of q reflectances S , we can form the q -by- 3 matrix SER_{new} . This leads to Eq. 4. For the unknown R_1 a least square solution is used, which leads to an improved estimate of the system's spectral transfer function.

$$P = SERR_1 \quad (4)$$

The goal of the second step is to calculate the reflectances of unknown color samples. We use the RGB-values from the digitized color samples p , the improved transfer function R_{new} and the spectral irradiance E in order to calculate spectral reflectances S (See Eq. 2).

Since $S(\lambda)$ varies smoothly for fragments we can accurately represent the spectral reflectance of a set of color standards with the first few components of a characteristic vector analysis [9]. In effect, this analysis allows us to reduce the dimensionality of S and leads to an algorithm that gives colorimetrically accurate spectral reflectance from red-green-blue output of the video digitizing system.

S_{mean} is defined as mean vector (1 by m) from the color checker reflectances at $m = 30$ equally-spaced wavelengths across the spectrum. S_{basis} (n by m matrix) denotes the characteristic vectors used. We use $n = 3$ characteristic vectors to represent the original data. A 1 -by- n vector of basis weights (denoted B) is calculated when solving Eq. 5 by inserting the digitized RGB values into p .

$$B = (p - S_{mean}ER)(S_{basis}ER)^{-1} \quad (5)$$

When we multiply S_{basis} by the appropriate vector B and add the result to S_{mean} , we can reconstruct any spectral reflectance S in our set of colors (Eq. 6). For a more detailed description of the algorithm see [3].

$$S = S_{mean} + BS_{basis} \quad (6)$$

The technique used is a method for examining a number of sets of multivariate response data and determining linear transformations of the data to a smaller number of parameters which contains essentially all the information in the original data.

3 Color estimation process

First, the three spectral variables - irradiance of the lightsource $E(\lambda)$, camera transfer function $R_k(\lambda)$ and reflectance $S(\lambda)$ of the MacBeth reference chart - have to be initialized.

We use Tungsten Halogen Floodlamps 7700 (150W) and TL-light as lightsources. In order to recover colorimetric data from our samples under a variety of lightsources we use different types of lightsources. The spectral distribution was given by the manufacturer. Figure 1 shows the typical spectral distribution of TL-82 and TL-95 with slight differences between these two lamps.

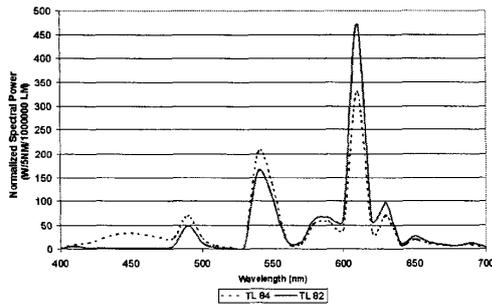


Figure 1. Spectral irradiance of TL-82 and TL-95

The video cameras used are a 3CCD DONPISHA XC-003P and a CCD-IKEGAMY ICD-700P. The Ikegamy camera is a single CCD-color CCTV camera, which is used to give out Y/C (chrominance/luminance) separation signals. The Sony camera is a color video module, which uses a CCD for the pick-up device. It has an RGB signal output. Both cameras are one-chip-cameras. Figure 2 shows the spectral response curve of the DONPISHA camera. The data was provided by the manufacturer.

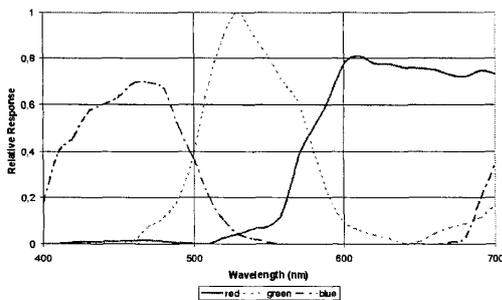


Figure 2. Typical spectral response of a Sony-camera

The spectral reflectance is scaled in equally-spaced wavelengths (every 10nm) across the spectrum. 12 colors of the MacBeth Color checker are used as a reference set

and 12 are used for evaluation purposes. Their reflectance is measured using a spectroradiometer. For our reference set we choose colors which have a similar spectral distribution to the colors of our archaeological findings in order to maximize the achievable accuracy of the vector analysis.

In the next step we grab an image of an archaeological fragment, which leads to RGB values. Test regions are specified manually, and their RGB-Values are used to reconstruct the reflectance. Figure 3 shows two different test regions A and B.

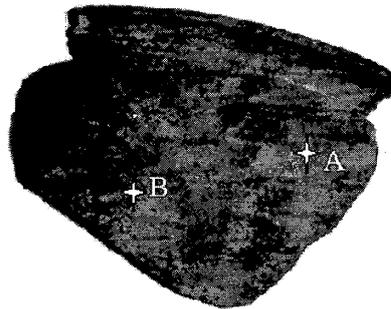


Figure 3. Test regions A and B

4 Results

Two experiments are presented: the first example with MacBeth Colors and the second with real fragments. In a first experiment we use the measured reflectance of 12 MacBeth color patches as reference and try to estimate the reflectance of the other 12 patches using the reference set. The resulting reflectance is compared to previous measured values.

Figure 4 shows the result for patch 1 (dark skin). In that case, the correlation equals 0,98. The computed reflectances of the other 11 patches correlated between and 0,85 and 0,98 to their corresponding measured reflectances with an average correlation of 0,92 (see Table 1). Lower correlation may be caused by the purely statistical representation of the underlying variables by the characteristic vector analysis.

In the second experiment we grab an image of a fragment and specify two test regions A and B (Figure 3). The reference set was chosen from the MacBeth color checker. The spectral reflectances of A and B are computed and visualized in Figure 5. For evaluation purposes we calculate CIE tristimulus values using a linear transformation and compare the achieved values with measured chromaticity coordinates.

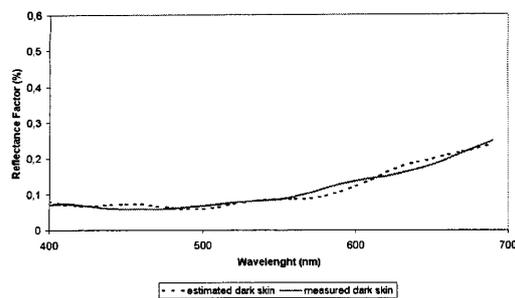


Figure 4. Measured and estimated spectral reflectance of a MacBeth Color Patch

patchnr	corr	patchnr	corr
1	0.98	7	0.97
2	0.97	8	0.95
3	0.93	9	0.96
4	0.98	10	0.91
5	0.86	11	0.85
6	0.92	12	0.89

Table 1. Correlation between measured and calculated spectral reflectances of 12 Macbeth ColorChecker patches

dinates from a Chroma Meter CR-200b. Table 2 shows a comparison between measured and computed chromaticity coordinates. The final results are in the close neighborhood of the measured values. Since these results are influenced by the linear transformation we plan measurements using a spectroradiometer in order to allow direct comparison between measured and computed reflectances.

	Comp. A	Meas. A	Comp. B	Meas.B
x	0.48	0.33	0.49	0.40
y	0.39	0.34	0.41	0.37
Y	17.9	11.1	32.3	21.0

Table 2. Measured and computed chromaticity coordinates

5 Conclusion and Outlook

In this work we presented a technique for accurate color estimation, which plays an important role in the classification process for archaeological fragments. We proposed an

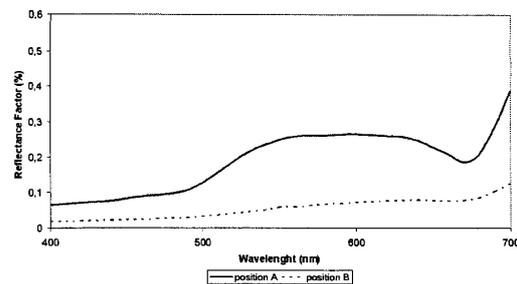


Figure 5. Calculated spectral reflectance of positions A and B

application using a straightforward approach based on a linear color calibration technique. Since the color specification of a fragment is gained by different archaeologists and under varying lightning conditions the results differ from each other. The results obtained give a good initial estimate to the archaeologists. Future work goes towards color calibration without known illuminants in order to allow color estimation outside laboratory conditions.

References

- [1] T. Gevers and W. M. Smeulders. Color-based Object Recognition. *Pattern Recognition*, 32(2):453–464, 1999.
- [2] M. Kampel and R. Sablatnig. On 3d Modelling of Archaeological Sherds. In *Proc. of Int. Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging*, pages 95–98, 1999.
- [3] R. L. Lee. Colorimetric calibration of a Video Digitizing System. *Colour Research and Application*, 13(3):180–186, 1988.
- [4] L. T. Maloney and B. A. Wandell. Colour Constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3(1):29–33, 1986.
- [5] C. S. McCamy, H. Marcus, and J. G. Davidson. A Colour-Rendition Chart. *Journal of Applied Photographic Engineering*, 2(3):95–99, 1976.
- [6] C. Menard and R. Sablatnig. Computer based Acquisition of Archaeological Finds: The First Step towards Automatic Classification. In H. Kamermans and K. Fennema, editors, *Interfacing the Past, Computer Applications and Quantitative Methods in Archaeology*, number 28, pages 413–424, Leiden, March 1996. *Analecta Praehistorica Leidensia*.
- [7] C. Menard and I. Tastl. Automated Color Determination for Archaeological Objects. In *Is&T 4th Color Imaging Conference*, pages 160–163, 1996.
- [8] S. Nayar and R. Bolle. Reflectance Based Object Recognition. *Int. Journal of Computer Vision*, 17(3):219–240, 1996.
- [9] J. L. Simonds. Application of characteristic vector analysis to photographic and optical response data. *Journal of the Optical Society of America*, 53:968–974, 1963.

Chapter 8

Improved Motion Segmentation based on Shadow Detection

Martin Kampel, Horst Wildenauer, Phlipp Blauensteiner and Allan Hanbury,
Improved Motion Segmentation based on Shadow Detection, *Electronic Letters on Computer Vision and Image Analysis*, vol 6(3), pp. 1-12, 2007.

Improved motion segmentation based on shadow detection

M. Kampel* and H. Wildenauer⁺ and P. Blauensteiner* and A. Hanbury*

* *Pattern Recognition and Image Processing Group, Vienna University of Technology, Favoritenstr.9, A-1040 Vienna, Austria*

⁺ *Automation and Control Institute, Vienna University of Technology, Gusshausstr.27, A-1040 Vienna, Austria*

Abstract

In this paper, we discuss common colour models for background subtraction and problems related to their utilisation are discussed. A novel approach to represent chrominance information more suitable for robust background modelling and shadow suppression is proposed. Our method relies on the ability to represent colours in terms of a 3D-polar coordinate system having saturation independent of the brightness function; specifically, we build upon an Improved Hue, Luminance, and Saturation space (IHLS). The additional peculiarity of the approach is that we deal with the problem of unstable hue values at low saturation by modelling the hue-saturation relationship using saturation-weighted hue statistics. The effectiveness of the proposed method is shown in an experimental comparison with approaches based on RGB, Normalised RGB and HSV.

Key Words: Motion detection, shadow detection, background subtraction, colour spaces.

1 Introduction

The underlying step of visual surveillance applications like target tracking and scene understanding is the detection of moving objects. Background subtraction algorithms are commonly applied to detect these objects of interest by the use of statistical colour background models. Many present systems exploit the properties of the Normalised RGB to achieve a certain degree of insensitivity with respect to changes in scene illumination.

Hong and Woo [1] apply the Normalised RGB space in their background segmentation system. McKenna et al. [2] use this colour space in addition to gradient information for their adaptive background subtraction. The AVITRACK project [3] utilises Normalised RGB for change detection and adopts the shadow detection proposed by Horprasert et al. [4].

Beside Normalised RGB, representations of the RGB colour space in terms of 3D-polar coordinates (hue, saturation, and brightness) are used for change detection and shadow suppression in surveillance applications. François and Medioni [5] suggest the application of HSV for background modelling for real-time video segmentation. In their work, a complex set of rules is introduced to reflect the relevance of observed and background colour information during change detection and model update. Cucchiara et al. [6] propose a RGB-based background model which they transform to the HSV representation in order to utilise the properties of HSV chrominance information for shadow suppression.

Correspondence to: <kampel@prip.tuwien.ac.at>

Recommended for acceptance by U. Pal and P. Nagabhushan

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

Our approach differs from the aforementioned in the way that we build upon the IHLS colour space, which is more suitable for background subtraction. Additionally, we propose the application of saturation-weighted hue statistics [7] to deal with unstable hue values at weakly saturated colours. Also, a technique to efficiently classify changes in scene illumination (e.g. shadows), modelling the relationship between saturation and hue has been devised.

The remainder of this paper is organised as follows: Section 2 reviews the Normalised RGB and the Improved Hue, Luminance and Saturation (IHLS) colour space. Furthermore it gives a short overview over circular colour statistics, which have to be applied on the hue as angular value. Section 3 presents how these statistics can be applied in order to model the background in image sequences. In Section 4 we describe metrics for the performance evaluation of our motion segmentation. The conducted experiments and their results are presented in Section 5. Section 6 concludes this paper and gives an outlook.

2 Colour Spaces

In this section, the Normalised RGB and IHLS colour spaces used in this paper are described. It also gives a short overview over circular colour statistics and a review of saturation weighted hue statistics.

2.1 Normalised RGB

The Normalised RGB space aims to separate the chromatic components from the brightness component. The red, green and blue channel can be transformed to their normalised counterpart by using the formulae

$$l = R + G + B, \quad r = R/l, \quad g = G/l, \quad b = B/l \quad (1)$$

if $l \neq 0$ and $r = g = b = 0$ otherwise [8]. One of these normalised channels is redundant, since by definition r , g , and b sum up to 1.

Therefore, the Normalised RGB space is sufficiently represented by two chromatic components (e.g. r and g) and a brightness component l . From Kender [9] it is known that the practical application of Normalised RGB suffers from a problem inherent to the normalisation; namely, that noise (such as, e.g. sensor or compression noise) at low intensities results in unstable chromatic components. For an example see Figure 1. Note the artefacts in dark regions such as the bushes (top left) and the shadowed areas of the cars (bottom right).



Figure 1: Examples of chromatic components. Lexicographically ordered - Image from the *PETS2001* dataset, it's normalised blue component b , normalised saturation (cylindrical HSV), IHLS saturation.

2.2 IHLS Space

The Improved Hue, Luminance and Saturation (IHLS) colour space was introduced in [10]. It is obtained by placing an *achromatic axis* through all the grey ($R = G = B$) points in the RGB colour cube, and then specifying the coordinates of each point in terms of position on the achromatic axis (brightness), distance from the axis (saturation s) and angle with respect to pure red (hue θ^H). The IHLS model is improved with respect to the similar colour spaces (HLS, HSI, HSV, etc.) by removing the normalisation of the saturation by the brightness. This has the following advantages: (a) the saturation of achromatic pixels is always low and (b) the saturation is independent of the brightness function used. One may therefore choose any function of R , G and B to calculate the brightness.

It is interesting that this normalisation of the saturation by the brightness, which results in the colour space having the shape of a cylinder instead of a cone or double-cone, is usually implicitly part of the transformation equations from RGB to a 3D-polar coordinate space. This is mentioned in one of the first papers on this type of transformation [11], but often in the literature the equations for a cylindrically-shaped space (i.e. with normalised saturation) are shown along with a diagram of a cone or double-cone (for example in [12, 13]). Figure 1 shows a comparison of the different formulations of saturation. The undesirable effects created by saturation normalisation are easily perceivable, as some dark, colourless regions (eg., the bushes and the side window of the driving car) reach higher saturation values than their more colourfull surroundings. Also, note the artefacts resulting from the singularity of the saturation at the black vertex of the RGB-cube (again, the bushes and the two bottom right cars).

The following formulae are used for the conversion from RGB to hue θ^H , luminance y and saturation s of the IHLS space:

$$\begin{aligned}
 s &= \max(R, G, B) - \min(R, G, B) \\
 y &= 0.2125R + 0.7154G + 0.0721B \\
 cr_x &= R - \frac{G+B}{2}, \quad cr_y = \frac{\sqrt{3}}{2}(B-G) \\
 cr &= \sqrt{cr_x^2 + cr_y^2} \\
 \theta^H &= \begin{cases} \text{undefined} & \text{if } cr = 0 \\ \arccos\left(\frac{cr_x}{cr}\right) & \text{elseif } cr_y \leq 0 \\ 360^\circ - \arccos\left(\frac{cr_x}{cr}\right) & \text{else} \end{cases}
 \end{aligned} \tag{2}$$

where cr_x and cr_y denote the chrominance coordinates and $cr \in [0, 1]$ the chroma. The saturation assumes values in the range $[0, 1]$ independent of the hue angle (the maximum saturation values are shown by the circle on the chromatic plane in Figure 2). The chroma has the maximum values shown by the dotted hexagon in Figure 2. When using this representation, it is important to remember that the hue is undefined if $s = 0$, and that it does not contain much useable information when s is low (i.e. near to the achromatic axis).

2.3 Hue statistics

In a 3D-polar coordinate space, standard (linear) statistical formulae can be utilised to calculate statistical descriptors for brightness and saturation coordinates. The hue, however, is an angular value, and consequently the appropriate methods from circular statistics are to be used.

Now, let θ_i^H , $i = 1, \dots, n$ be n observations sampled from a population of angular hue values. Then, the vector \mathbf{h}_i pointing from $\mathbf{O} = (0, 0)^T$ to the point on the circumference of the unit circle, corresponding to θ_i^H , is given by the Cartesian coordinates $(\cos \theta_i^H, \sin \theta_i^H)^T$.*

*Note that, when using the IHLS space (Eq. 3), no costly trigonometric functions are involved in the calculation of \mathbf{h}_i , since $\cos(\theta_i^H) = cr_x/cr$ and $\sin(\theta_i^H) = -cr_y/cr$.

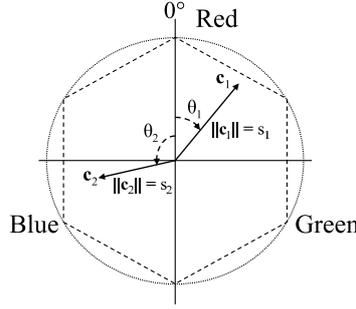


Figure 2: The chromatic plane of the IHS color space.

The mean direction $\bar{\theta}^H$ is defined to be the direction of the resultant of the unit vectors $\mathbf{h}_1, \dots, \mathbf{h}_n$ having directions θ_i^H . That is, we have

$$\bar{\theta}^H = \arctan2(\mathcal{S}, \mathcal{C}), \quad (3)$$

where

$$\mathcal{C} = \sum_{i=1}^n \cos \theta_i^H, \quad \mathcal{S} = \sum_{i=1}^n \sin \theta_i^H \quad (4)$$

and $\arctan2(y, x)$ is the four-quadrant inverse tangent function.

The mean length of the resultant vector

$$\bar{\mathcal{R}} = \frac{\sqrt{\mathcal{C}^2 + \mathcal{S}^2}}{n}. \quad (5)$$

is an indicator of the dispersion of the observed data. If the n observed directions θ_i^H cluster tightly about the mean direction $\bar{\theta}^H$ then $\bar{\mathcal{R}}$ will approach 1. Conversely, if the angular values are widely dispersed $\bar{\mathcal{R}}$ will be close to 0. The *circular variance* is defined as

$$\mathcal{V} = 1 - \bar{\mathcal{R}} \quad (6)$$

While the circular variance differs from the linear statistical variance in being limited to the range $[0, 1]$, it is similar in the way that lower values represent less dispersed data. Further measures of circular data distribution are given in [14].

2.4 Saturation-weighted hue statistics

The use of statistics solely based on the hue has the disadvantage of ignoring the tight relationship between the chrominance components hue and saturation. For weakly saturated colours the hue channel is unimportant and behaves unpredictably in the presence of colour changes induced by image noise. In fact, for colours with zero saturation the hue is undefined.

As one can see in Figure 2, the chromatic components may be represented by means of Cartesian coordinate vectors \mathbf{c}_i with direction and length given by hue and saturation respectively. Using this natural approach, we introduce the aforementioned relationship into the hue statistics by weighting the unit hue vectors \mathbf{h}_i by their corresponding saturations s_i .

Now, let $(\theta_i^H, s_i), i = 1, \dots, n$ be n pairs of observations sampled from a population of hue values and associated saturation values. We proceed as described in Section 2.3, with the difference that instead of calculating the resultant of unit vectors, the vectors \mathbf{c}_i , which we will dub *chrominance vectors* throughout this paper, have length s_i .

That is, we weight the vector components in Eq. 4 by their saturations s_i

$$\mathcal{C}_s = \sum_{i=1}^n s_i \cos \theta_i^H, \quad \mathcal{S}_s = \sum_{i=1}^n s_i \sin \theta_i^H, \quad (7)$$

and choose the mean resultant length of the chrominance vectors (for other possible formulations see, e.g. [7]) to be

$$\overline{\mathcal{R}}_n = \frac{\sqrt{\mathcal{C}_s^2 + \mathcal{S}_s^2}}{n}. \quad (8)$$

Consequently, for the mean resultant chrominance vector we get

$$\overline{\mathbf{c}}_n = (\mathcal{C}_s/n, \mathcal{S}_s/n)^T. \quad (9)$$

Here, the length of the resultant is compared to the length obtained if all vectors had the same direction and maximum saturation. Hence, $\overline{\mathcal{R}}_n$ gives an indication of the saturations of the vectors which gave rise to the mean of the chrominance vector, as well as an indication of the angular dispersion of the vectors. To test if a mean chrominance vector $\overline{\mathbf{c}}_n$ is similar to a newly observed chrominance vector, we use the Euclidean distance in the chromatic plane:

$$D = \sqrt{(\overline{\mathbf{c}}_n - \mathbf{c}_o)^T (\overline{\mathbf{c}}_n - \mathbf{c}_o)}, \quad (10)$$

with $\mathbf{c}_o = s_o \mathbf{h}_o$. Here, \mathbf{h}_o and s_o denote the observed hue vector and saturation respectively.

3 The IHLS Background Model

With the foundations laid out in Section 2.4 we proceed with devising a simple background subtraction algorithm based on the IHLS colour model and saturation-weighted hue statistics. Specifically, each background pixel is modelled by its mean luminance μ_y and associated standard deviation σ_y , together with the mean chrominance vector $\overline{\mathbf{c}}_n$ and the mean Euclidean distance σ_D between $\overline{\mathbf{c}}_n$ and the observed chrominance vectors (see Eq. 10).

On observing the luminance y_o , saturation s_o , and a Cartesian hue vector \mathbf{h}_o for each pixel in a newly acquired image, the pixel is classified as foreground if:

$$|(y_o - \mu_y)| > \alpha \sigma_y \quad \vee \quad \|\overline{\mathbf{c}}_n - s_o \mathbf{h}_o\| > \alpha \sigma_D \quad (11)$$

where α is the foreground threshold, usually set between 2 and 3.5.

In order to decide whether a foreground detection was caused by a moving object or by its shadow cast on the static background, we exploit the chrominance information of the IHLS space. A foreground pixel is considered as shaded background if the following three conditions hold:

$$y_o < \mu_y \quad \wedge \quad |y_o - \mu_y| < \beta \mu_y, \quad (12)$$

$$s_o - \overline{\mathcal{R}}_n < \tau_{ds} \quad (13)$$

$$\|\mathbf{h}_o \overline{\mathcal{R}}_n - \overline{\mathbf{c}}_n\| < \tau_h, \quad (14)$$

where $\overline{\mathcal{R}}_n = \|\overline{\mathbf{c}}_n\|$ (see Eq. 8).

These equations are designed to reflect the empirical observations that cast shadows cause a darkening of the background and usually lower the saturation of a pixel, while having only limited influence on its hue. The first condition (Eq. 12) works on the luminance component, using a threshold β to take into account the strength of the predominant light source. Eq. 13 performs a test for a lowering in saturation, as proposed by Cucchiara et al. [6]. Finally, the lowering in saturation is compensated by scaling the observed hue vector \mathbf{h}_o to the same length as the mean chrominance vector $\overline{\mathbf{c}}_n$ and the hue deviation is tested using the Euclidean distance (Eq. 14).

This, in comparison to a check of angular deviation (see Eq. 31 or [6]), also takes into account the model's confidence in the learned chrominance vector. That is, using a fixed threshold τ_h on the Euclidean distance relaxes the angular error-bound in favour of stronger hue deviations at lower model saturation value $\overline{\mathcal{R}}_n$, while penalising hue deviations for high saturations (where the hue is usually more stable).

4 Metrics for Motion Segmentation

The quality of motion segmentation can in principle be described by two characteristics. Namely, the spatial deviation from the reference segmentation, and the fluctuation of spatial deviation over time. In this work, however, we concentrate on the evaluation of spatial segmentation characteristics. That is, we will investigate the capability of the error metrics listed below, to describe the spatial accuracy of motion segmentations.

- Detection rate (DR) and false alarm rate (FR)

$$DR = \frac{TP}{FN + TP} \quad (15)$$

$$FR = \frac{FP}{N - (FN + TP)} \quad (16)$$

where TP denotes the number of true positives, FN the number of false negatives, FP the number of false positives, and N the total number of pixels in the image.

- Misclassification penalty (MP)

The obtained segmentation is compared to the reference mask on an object-by-object basis; misclassified pixels are penalized by their distances from the reference objects border [15].

$$MP = MP_{fn} + MP_{fp} \quad (17)$$

with

$$MP_{fn} = \frac{\sum_{j=1}^{N_{fn}} d_{fn}^j}{D} \quad (18)$$

$$MP_{fp} = \frac{\sum_{k=1}^{N_{fp}} d_{fp}^k}{D} \quad (19)$$

Here, d_{fn}^j and d_{fp}^k stand for the distances of the j^{th} false negative and k^{th} false positive pixel from the contour of the reference segmentation. The normalised factor D is the sum of all pixel-to-contour distances in a frame.

- Rate of misclassifications (RM)

The average normalised distance of detection errors from the contour of a reference object is calculated using [16]:

$$RM = RM_{fn} + RM_{fp} \quad (20)$$

with

$$RM_{fn} = \frac{1}{N_{fn}} \sum_{j=1}^{N_{fn}} \frac{d_{fn}^j}{D_{diag}} \quad (21)$$

$$RM_{fp} = \frac{1}{N_{fp}} \sum_{k=1}^{N_{fp}} \frac{d_{fp}^k}{D_{diag}} \quad (22)$$

N_{fn} and N_{fp} denote the number of false negative and false positive pixels respectively. q D_{diag} is the diagonal distance within the frame.

- Weighted quality measure (*QMS*)

This measure quantifies the spatial discrepancy between estimated and reference segmentation as the sum of weighted effects of false positive and false negative pixels [17].

$$QMS = QMS_{fn} + QMS_{fp} \quad (23)$$

with

$$QMS_{fn} = \frac{1}{N} \sum_{j=1}^{N_{fn}} w_{fn}(d_{fn}^j) d_{fn}^j \quad (24)$$

$$QMS_{fp} = \frac{1}{N} \sum_{k=1}^{N_{fp}} w_{fp}(d_{fp}^k) d_{fp}^k \quad (25)$$

N is the area of the reference object in pixels. Following the argument that the visual importance of false positives and false negatives is not the same, and thus they should be treated differently, the weighting functions w_{fp} and w_{fn} were introduced:

$$w_{fp}(d_{fp}) = B_1 + \frac{B_2}{d_{fp} + B_3} \quad (26)$$

$$w_{fn}(d_{fn}) = C \cdot d_{fn} \quad (27)$$

In our work for a fair comparison of the change detection algorithms with regard to their various decision parameters, receiver operating characteristics (ROC) based on detection rate (DR) and false alarm rate (FR) were utilised.

5 Experiments and Results

We compared the proposed IHLS method with three different approaches from literature. Namely, a RGB background model using either NRGB- (*RGB+NRGB*), or HSV-based (*RGB+HSV*) shadow detection, and a method relying on NRGB for both background modelling and shadow detection (*NRGB+NRGB*).

All methods were implemented using the *Colour Mean and Variance* approach to model the background [18]. A pixel is considered foreground if $|c_o - \mu_c| > \alpha \sigma_c$ for any channel c , where $c \in \{r, g, l\}$ for the Normalised RGB and $c \in \{R, G, B\}$ for the RGB space respectively. o_c denotes the observed value, μ_c its mean, σ_c the standard deviation, and α the foreground threshold.

The tested background models are maintained by means of exponentially weighted averaging [18] using different learning rates for background and foreground pixels. During the experiments the same learning and update parameters were used for all background models, as well as the same number of training frames.

For Normalised RGB (*RGB+NRGB*, *NRGB+NRGB*), shadow suppression was implemented based on Horprasert's approach [3, 4]. Each foreground pixel is classified as shadow if:

$$\begin{aligned} l_o < \mu_l \quad \wedge \quad l_o > \beta \mu_l \\ |r_o - \mu_r| < \tau_c \quad \wedge \quad |g_o - \mu_g| < \tau_c \end{aligned} \quad (28)$$

where β and τ_c denote thresholds for the maximum allowable change in the intensity and colour channels, so that a pixel is considered as shaded background.

In the HSV-based approach (*RGB+HSV*) the RGB background model is converted into HSV (specifically, the reference luminance μ_v , saturation μ_s , and hue μ_θ) before the following shadow tests are applied. A foreground pixel is classified as shadow if:

$$\beta_1 \leq \frac{v_o}{\mu_v} \leq \beta_2 \quad (29)$$

$$s_o - \mu_s \leq \tau_s \quad (30)$$

$$|\theta_o^H - \mu_\theta| \leq \tau_\theta \quad (31)$$

The first condition tests the observed luminance v_o for a significant darkening in the range defined by β_1 and β_2 . On the saturation s_o a threshold on the difference is performed. Shadow lowers the saturation of points and the difference between images and the reference is usually negative for shadow points. The last condition takes into account the assumption that shading causes only small deviation of the hue θ_o^H [6].

For the evaluation of the algorithms, three video sequences were used. As an example for a typical indoor scene *Test Sequence 1*, recorded by an AXIS-211 network camera, shows a moving person in a stairway. For this sequence, ground truth was generated manually for 35 frames. *Test Sequence 2* was recorded with the same equipment and shows a person waving books in front of a coloured background. For this sequence 20 ground truth frames were provided. Furthermore in *Test Sequence 3* the approaches were tested on 25 ground truth frames from the *PETS2001* dataset 1 (camera 2, testing sequence). Example pictures of the dataset can be found in Figure 3.

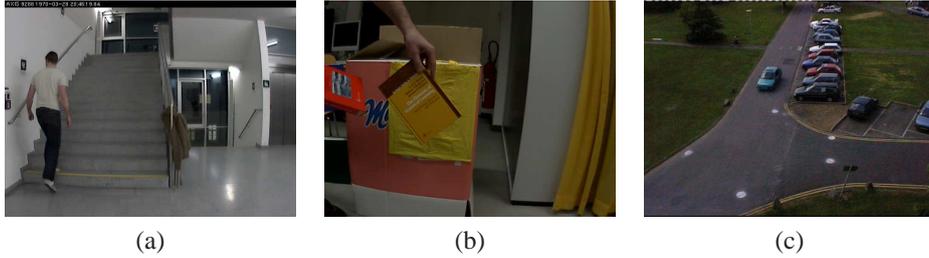


Figure 3: Evaluation dataset: *Test Sequence 1* (a), *Test Sequence 2* (b), *Test Sequence 3* (c)

For a dense evaluation, we experimentally determined suitable ranges for all parameters and sub-sampled them in ten steps. Figure 7 shows the convex hulls of the points in ROC space obtained for all parameter combinations. We also want to point out that *RGB+HSV* was tested with unnormalised and normalised saturation; however, since the normalised saturation consistently performed worse, we omit the results in the ROC for clarity of presentation.

As one can see, our approach outperforms its competitors on *Test Sequence 1*. One reason for this is the insensitivity of the *RGB+NRGB* and *NRGB+NRGB* w.r.t. small colour differences at light, weakly saturated colours. *RGB+HSV*, however, suffered from the angular hue test reacting strongly to unstable hue values close to the achromatic axis. For conservative thresholds (i.e. small values for τ_c or τ_θ) all three approaches either detected shadows on the wall as foreground, or, for larger thresholds failed to classify the beige t-shirt of the person as foreground. Figure 4 shows output images from *Test Sequence 1*. We present the source image (a), the ground truth image (b), the resulting image from our approach (c), and the resulting images from the algorithms we compared with. I.a. it is shown that the shirt of the person in image (c) is detected with higher precision as in the images (d), (e), and (f), where it is mostly marked as shadow.

For *Test Sequence 2* the advantageous behaviour of our approach is even more evident. Although the scene is composed of highly saturated, stable colours, *RGB+NRGB* and *NRGB+NRGB* show rather poor results, again stemming from their insufficient sensitivity for bright colours. *RGB+HSV* gave better results, but could not take full advantage of the colour information. Similar hue values for the books and the background resulted

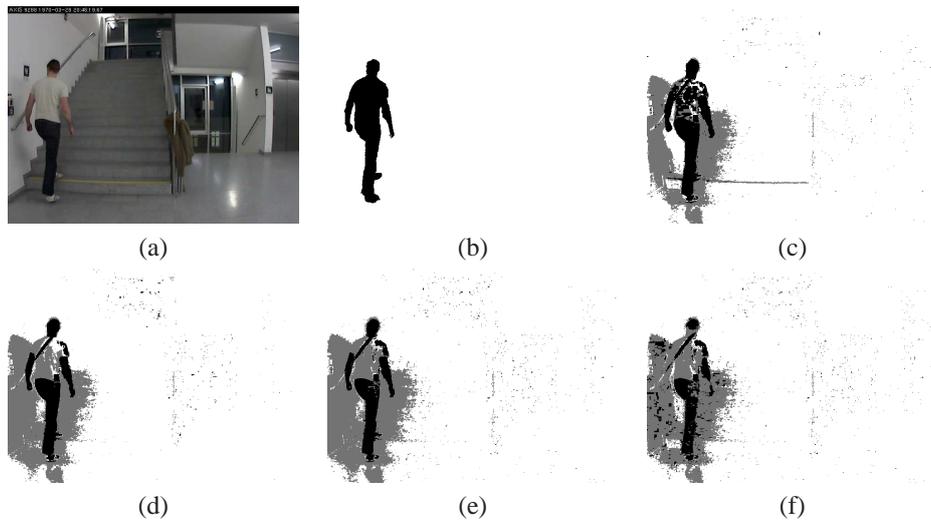


Figure 4: Output images *Test Sequence 1*: *Source Image* (a), *Ground Truth* (b), *Our Approach* (c), *RGB+NRGB* (d), *NRGB+NRGB* (e), *RGB+HSV* (f)

in incorrectly classified shadow regions. Figure 5 shows output images from *Test Sequence 2*. Especially the lower left part of the images (c), (d), (e), and (f) visualizes a better performance of the IHLS approach.

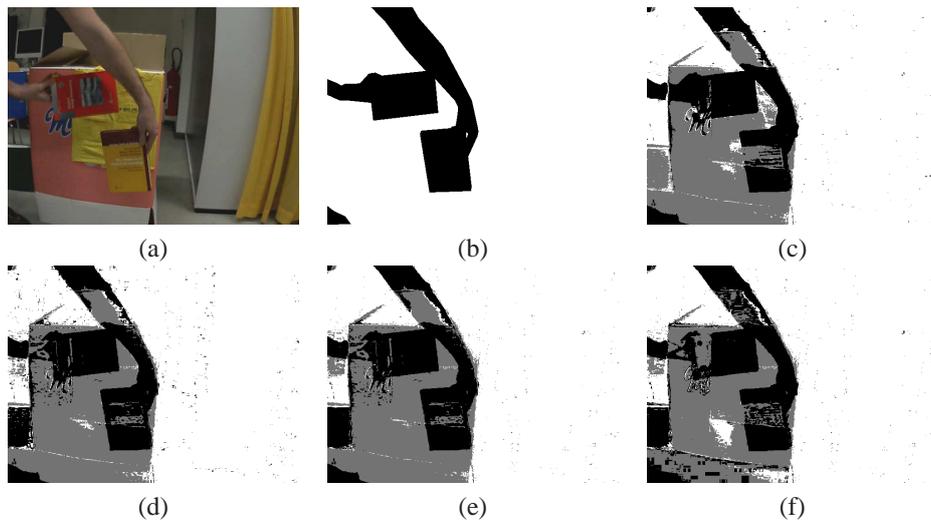


Figure 5: Output images *Test Sequence 2*: *Source Image* (a), *Ground Truth* (b), *Our Approach* (c), *RGB+NRGB* (d), *NRGB+NRGB* (e), *RGB+HSV* (f)

The *Test Sequence 3* sequence shows the problems of background modelling using NRGB already mentioned in Section 2. Due to the low brightness and the presence of noise in this scene, the chromatic components are unstable and therefore the motion detection resulted in an significantly increased number of false positives. *RGB+NRGB* and our approach exhibit similar performance (our approach having the slight edge), mostly relying on brightness checks, since there was not much useable information in shadow regions. *RGB+HSV* performed less well, having problems to cope with the unstable hue information in dark areas. Figure 6 shows output images *Test Sequence 3*.

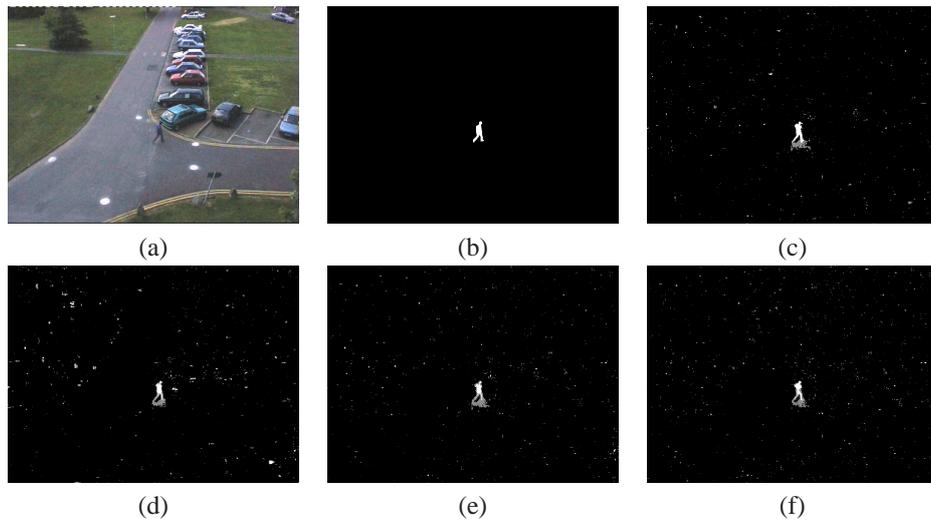


Figure 6: Output images *Test Sequence 3*: *Source Image* (a), *Ground Truth* (b), *Our Approach* (c), *RGB+NRGB* (d), *NRGB+NRGB* (e), *RGB+HSV* (f)

6 Conclusion

We proposed the usage of the IHLS colour space for change detection and shadow suppression in visual surveillance tasks. In the proposed framework, we advocate the application of saturation-weighted hue statistics to deal with the problem of the unstable hue channel at weakly saturated colours.

We have shown that our approach outperforms the approaches using Normalised RGB or HSV in several challenging sequences. Furthermore, our experiments have shown that it is not advisable to use NRGB for background modelling due to its unstable behaviour in dark areas.

One problem of our approach, however, is the fact that due to the use of saturation weighted hue statistics, it is impossible to tell whether a short chrominance vector in the background model is the result of unstable hue information or of a permanent low saturation. Although in the conducted experiments no impairments were evident, it is subject of further research in which cases this shortcoming poses a problem. Other fields of interest are the examination of alternatives to the Euclidean distance for the comparison of the chrominance vectors and an experimental in-depth-investigation of the shadow classification.

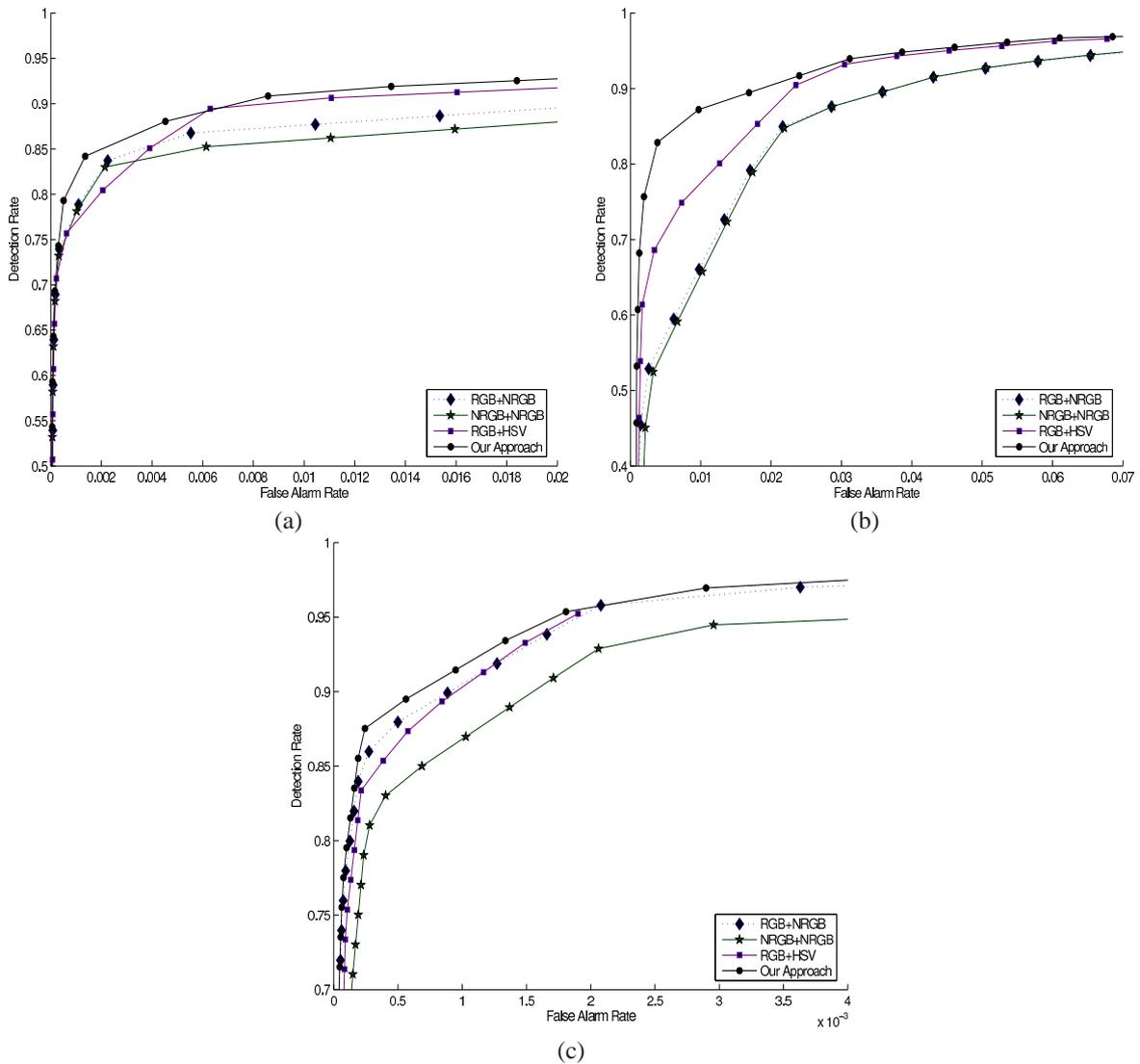


Figure 7: Experimental results: ROCs for *Test Sequence 1* (a), *Test Sequence 2* (b), and *Test Sequence 3* (c).

References

- [1] D. Hong & W. Woo, A Background Subtraction for Vision-based User Interface, *Proc. 4th International Conference on Information, Communications and Signal Processing and Fourth IEEE Pacific-Rim Conference on Multimedia*, 2003, 263–267.
- [2] S. McKenna, S. Jabri, Z. Duric, H. Wechsler, & A. Rosenfeld, Tracking Groups of People, *Computer Vision and Image Understanding*, 80(1), 2000, 42–56.
- [3] J. Ferryman, M. Borg, D. Thirde, F. Fusier, V. Valentin, F. Br mond, M. Thonnat, J. Aguilera, & M. Kampel, Automated Scene Understanding for Airport Aprons, *Proc. Australian Joint Conference on Artificial Intelligence*, Australia, 2005, 593–503.

- [4] T. Horprasert, D. Harwood, & L. Davis, A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection, *Proc. IEEE Conference on Computer Vision, FRAME-RATE Workshop*, 1999.
- [5] A. R. J. François & G. G. Medioni, Adaptive Color Background Modeling for Real-Time Segmentation of Video Streams, *Proc. International Conference on Imaging Science, Systems, and Technology*, 1999, 227–232.
- [6] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, & S. Sirotti, Improving Shadow Suppression in Moving Object Detection with HSV Color Information, *Proc. Intelligent Transport Systems*, 2001, 334–339.
- [7] A. Hanbury & W. G. Kropatsch, Colour Statistics for Matching in Image Databases, *Proc. 27th OEAGM Workshop*, Austria, 2003, 221–228.
- [8] G. D. Finlayson, B. Schiele, & J. L. Crowley, Comprehensive Colour Image Normalization, *Proc. 5th European Conference on Computer Vision*, 1998, 475–490.
- [9] J. R. Kender, Saturation, Hue and Normalized Colors: Calculation, Digitisation Effects, and Use, Tech. rep., Department of Computer Science, Carnegie Mellon University, 1976.
- [10] A. Hanbury, A 3D-polar coordinate colour representation well adapted to image analysis, *Proc. 13th Scandinavian Conference on Image Analysis*, 2003, 804–811.
- [11] A. R. Smith, Color gamut transform pairs, *Computer Graphics*, 12(3), 1978, 12–19.
- [12] R. C. Gonzalez & R. E. Woods, *Digital Image Processing* (Englewood Cliffs, NJ: Prentice-Hall, 2002).
- [13] D. Hearn & M. P. Baker, *Computer Graphics* (Englewood Cliffs, NJ: Prentice Hall, 1997).
- [14] K. V. Mardia, *Statistics of Directional Data* (London, UK: Academic Press, 1972).
- [15] C. Erdem & B. Sankur, Performance Evaluation Metrics for Object-Based Video Segmentation, *Proc. 10th European Signal Processing Conference, Tampere, Finland*, 2000, 917–920.
- [16] T. Schlögl, C. Beleznai, M. Winter, & H. Bischof, Performance Evaluation Metrics for Motion Detection and Tracking, *Proc. Proceedings of the International Conference on Pattern Recognition, Cambridge, UK*, 2004, 519–522.
- [17] P. Villegas & X. Marichal, Perceptually-Weighted Evaluation Criteria for Segmentation Masks in Video Sequences, *IEEE Transactions on Image Processing*, (8), 2004, 1092–1103.
- [18] C. R. Wren, A. Azarbayejami, T. Darrel, & A. Pentland, Pfunder: Real-Time Tracking of the Human Body, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 1997, 780–785.