

PRIP-TR-80

May 19, 2003

Shape Based Machine Vision¹

Robert Sablatnig

Abstract

The study of visual object recognition is often motivated by the problem of recognizing 3-d objects given that we receive 2-d patterns of light on our retinae. Recent findings from human psychophysics, neurophysiology and computational vision provide converging evidence for a view-based recognition framework in which objects and scenes are represented as collections of viewpoint-specific local features rather than 2-d templates or 3-d models. Hence the recent decade saw a gradual shift away from the 3-d object reconstruction approach pioneered by Marr toward view-based approaches. This report summarizes our contributions to this problem where we focus on the shape as recognition feature and apply these findings in the area of Machine Vision. The first part presents an overview of the framework, motivates the view-based recognition strategy, and introduces the hierarchical matching concept. Next, a short summary of a collection of six representative publications of our work carried out in this field, and a discussion of how this fits into the framework is given. The second part consists of the six papers themselves, where we start with a paper on the general framework which is followed by three different applications of the framework in Visual Inspection, Archaeology and Art History. The remaining two papers describe recent work performed in 3-d vision as part of the object-based recognition concept. The first paper is on the registration of range data, in which we propose a novel technique for range image registration. The collection ends with a work on combining different 3-d acquisition techniques within the hierarchical framework.

¹ 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.

Contents

1	Overview	2
1.1	Machine Vision and Shape Definition	3
1.2	View Based Recognition	6
1.3	Hierarchical Matching	8
1.4	Summary of Papers	13
1.5	Discussion of the Papers	17
	References	23
2	Increasing Flexibility for Automatic Visual Inspection: The General Analysis Graph (“GANAG Framework”)	35
3	Application Constraints in the Design of an Automatic Reading Device for Analog Display Instruments (“Inspection”)	47
4	Automated Segmentation of Archaeological Profiles for Classification (“Archaeology”)	56
5	Structural Analysis of Paintings Based on Brush Strokes (“Art History”)	61
6	Model-based Registration of Front- and Backviews (“Registration”)	74
7	Combining Shape from Silhouette and Shape from Structured Light for Volume Estimation of Archaeological Vessels (“Combination”)	91

Chapter 1

Overview

In nature, vision systems are of paramount importance to survival. The eye-brain combination makes it possible to absorb, process and react to large amounts of information about the surroundings, all without any physical contact [117]. The human visual system as a functional unit including the eyes, the nervous system, and the corresponding parts of the brain certainly ranks among the most important means of information processing. Typical vision tasks that humans perform nearly without any conscious effort are:

- recognition of “interesting” details in a complex scene (e.g. a good friend on a busy street);
- fast interpretation of local changes and appropriate reaction (e.g. driving a car);
- visual comparison (e.g. identification of a known human face);
- storing and retrieving of pictures (e.g. the local environment where one lives, a mountain scenery, etc.);

The efficiency of the biological systems in such areas are beyond the capabilities of today’s technical systems even with the fastest available computer systems. Nevertheless, it has now been well over 30 years since several individuals and groups made concerted efforts to automate visual perception in the research discipline of Computer Vision [65]. Computer Vision, sometimes also called image understanding or scene analysis, describes the automatic deduction of the structure and properties of a possibly dynamic three-dimensional (3-d) world from either a single or multiple two-dimensional (2-d) images of the world as a combination of image processing, pattern recognition, and artificial intelligence technologies [28]. Computer Vision describes a process that tries to recognize and locate position

and orientation, as well as to describe imaged objects in a 3-d environment like the human visual system does. An *object* is globally defined as something mental or physical toward which thought, feeling, or action is directed. In our restricted view for human vision, an object is something tangible and towards which an action can be directed. Therefore, subsequent discussions will not treat spatiotemporal variations of amorphous material distributions with diffuse borders such as clouds of smoke, objects are supposed to have a (at least visually) well defined surface.

Since the description of the state of the physical world from inherently noisy and ambiguous images of the world is a complicated goal to be accomplished in a reliable, robust, and efficient manner, there have been attempts to solve this problem in its full generality; however, there is no common solution available.

This habilitation thesis is a collection of different papers, each describing a vision or recognition task for a specific type of application. Let us first have a closer look into Machine Vision and Shape properties, then shortly summarize the view-based recognition approach (Section 1.2) and which strategies can be used to perform the matching in recognition (Section 1.3). Section 1.4 briefly summarizes the papers and Section 1.5 discusses how they fit into the framework presented.

1.1 Machine Vision and Shape Definition

From the early stages in computer vision researchers tried to convert the results achieved in basic research into applications to prove that their algorithms work. The pattern recognition and more generally the computer vision field has the potential and promise to provide the technology to develop a variety of automated systems that are capable of operating under diverse conditions, delivering consistent results, working in environments not suitable for humans, and situations where human workers have to perform a repetitive, tiring and error prone task [11]. Many different (often called "real-world") applications of computer vision were introduced, making work easier in the fields of navigation, manufacturing, quality control, remote sensing, cartography, target recognition and tracking, medical image analysis, document analysis, archaeology and art, to name a few. Generally, these applications are multidisciplinary and require a combination of science, engineering, and art, resulting in a challenging task to develop a successful computer vision solution [11].

Historically, the first industrial application area was in the manufacturing industry, because there was a strong desire to automate the production process and to control the final product. This application of computer vision is therefore called machine vision; the relationship between machine automation and computer vision is also represented in the

choice of the term. The term was introduced in the mid 80s; Batchelor, one of the pioneers in this field, first used the expression "practical pattern recognition" [7] and changed it to Machine Vision in 1985 [9]. Haralick defines a machine vision system as "a system capable of acquiring one or more images of an object, capable of processing, analyzing and measuring various characteristics of the acquired images, and interpreting the results of measuring in such a way that some useful decision can be made about the object" [28]. This global definition shows that there are various different applications in industry in general, however, there is no unique definition that can be found in the literature (see for instance [9, 8, 28, 11, 28, 60, 117, 88, 81]).

Figure 1.1 shows the hierarchy of vision terms in a simplified diagram, all terms are embedded in pattern recognition, since all use its methods. Other image analysis tasks are not mentioned explicitly, other computer vision research areas are shown in an exemplary manner only. There is no clear boundary between the tasks, the proposed subdivision is structured hierarchically - from low machine vision tasks, like locating, to more sophisticated tasks like guidance and control. Motion estimation applications for instance may also use recognition and identification techniques, measurement operations will often result in an accept/reject decision, which can be considered as inspection. However, there are applications that use only a limited number of pattern recognition techniques.

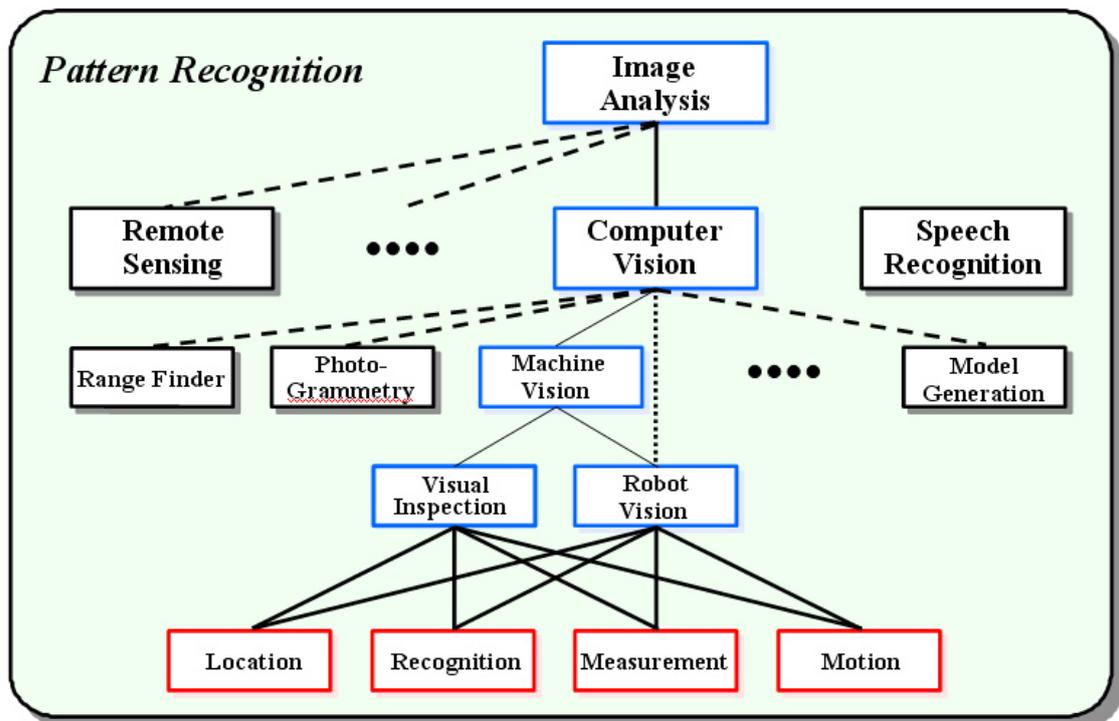


Figure 1.1: Hierarchical taxonomy of vision terms (simplified).

A priori knowledge about the object is used implicitly or explicitly by all machine vision systems, visual inspection for example can only be performed by matching the object under inspection with a set of predefined conditions of acceptability. The a priori known specifications are described by an explicit object model (a system of assumptions, data, and inferences, that describe the object), where all relevant object features are described. A typical method to build the object model consists of defining the radiometric and geometric features of the object [91]. Since the radiometric features of an object vary under different illumination conditions, geometric features, also called *shape features*, are preferred (shape features however also vary due to different viewing directions).

This work refers to shape based machine vision, so it is necessary to consider what is meant by "shape". Most people undoubtedly have an intuitive feeling for the meaning of the term, sensing its relatedness to such concepts as form and structure. Nevertheless, a precise definition of shape has proved elusive. Instead, researchers have adopted the working definition that shape is an aspect of a stimulus that remains invariant despite changes in size, position, and orientation [106]. For example, 2-d visual stimuli have the same shape if there exists a transformation of spatial scale (e.g., magnification) or a rotation in the picture plane that renders them identical [107]. Similarly, 3-d objects have the same shape if their volumes can be equated by size changes or a combination of rotations about three spatial axes.

In our view the term *shape* is used for the description of the geometric form of the object generated by its surfaces. The object can be characterized in this sense by a set of surface normals (shape) and a set of intensity values like surface texture, color, and reflection. The shape geometry can be characterized at any time as a topologically connected set of points, which is topologically compact (closed and bounded) and which is invariant to Euclidian geometric mappings [49, 54]. This means that *shape* in three dimensions is a volume or, more generally a surface. Note that every volume is bounded by a surface, but not every surface bounds a volume, the surface may not even be extendible to another surface that encloses the volume (for example a Möbius strip [27]). The reconstruction of three-dimensional object surfaces based on two dimensional visual mappings is only possible for visible object surfaces [22, 114]. An object is entirely visible from outside, if there exists a ray to each point of the surface that intersects the object only at this point.

Apart from our shape definition, other shape definitions also play an important role in the *Gestalt theory* of psychology [50, 48, 72] (for instance visual grouping of point patterns) and the *shape theory* of mathematical analysis [70] (for instance characterization of mathematical surfaces in different dimensions).

1.2 View Based Recognition

The study of visual object recognition is motivated by the problem of recognizing 3-d objects given that we only receive 2-d patterns of light on our sensor. Marr stated that the goal of vision is to reconstruct the scene [58]. Reconstruction assumes that visual perception is a hierarchical process which starts with local features that are combined into more complex descriptions [71]. At the endpoint of the reconstruction process Marr assumed that viewer- centered descriptions (sketches) are remapped into 3-d object centered representations [59]. This back-mapping is necessary to have a stable and invariant (over changes in the sensor) representation of the image. Object representations should therefore be object-centered rather than viewer-centered. Based on this theory machine vision researchers tried to implement reconstruction algorithms with only marginal success [65]. Thus, one strong argument for the view-based¹ approach is that it does not require reconstruction. This follows also the human vision theory since there viewer-centered images are also taken as input for recognition [111].

Critics of the reconstructionist school favor in contrast the theory of Purposive Vision, a paradigm also known as Behavioral Vision. In simple terms, this approach suggests that vision has a purpose, a goal. Often this goal is action; it can be theoretical, practical or aesthetic. When vision is considered in conjunction with action, it becomes easier. The reason is that the descriptions of space-time, that the system needs to derive, are not general purpose, but are purposive [4]. Thus, vision is more readily understood in the context of the behaviors in which the system is engaged [5]. Consequently, vision attempts to explore the aspects of the world that are important for the system at a given point in time, instead of aiming at a general representation of the environment which, besides being extremely difficult to extract, is probably not needed either. The interest in purposive vision is largely motivated by the fact that all biological vision systems are highly active and purposive [14]. The purposiveness of visual processes enables the formulation and the solution of simpler problems that have a relatively small number of possible solutions and can be treated in a qualitative manner[3]. This means that these descriptions are good for restricted sets of tasks, such as tasks related to navigation, manipulation and recognition. In our view, however, global recognition of objects does not always have a specific purpose.

Object recognition and classification research are both concerned with the question "what is the object?". To recognize an object as a car is not very different to putting the object into the class "car". Thus both object classification and object recognition research have to solve common problems, although these two research areas have evolved separately [104]. One reason for this is the difference of focus of these two research areas:

¹We use the term view-based, other common terms for this theory are exemplar-based, appearance-based, and image-based

typical classification studies have investigated the rules for the formation of classes, while recognition research has mostly explored the perceptual characteristics of the recognition process [6]. However, recent developments show that the principals governing the formation of classes are now coupled with the perceptual aspects of recognition [105]. Therefore, classification and recognition are closely related research areas which enable us to use a common research base and see the classification task as a more general recognition task (if we want to classify an object as car or vehicle we can also recognize it as a Ferrari or as the Ferrari of a specific person).

The image of an object changes as a function of viewpoint, lighting, size or location, but we are nevertheless able to interpret these images correctly. In Figure 1.2 one can recognize all the different chairs without difficulty despite the fact that the images of the chairs differ in shape and scale [16]. One can also distinguish easily between cast shadows and paintings of chairs although they have exactly the same shape. Furthermore, it is clear that the chair on the desk is a model chair that one can hold in the hand, whereas the chair in the next room is large enough to sit on, although they have exactly the same size in the image. Thus the information of the surroundings is essential to interpret the scene correctly.



Figure 1.2: A set of chairs that differ in size and shape (from [16]).

From the example given, it is evident that images from the same objects are highly variable, depending on the point of view and the acquisition technique used to get these representations [108]. Human object recognition experiments showed that if two views of unfamiliar objects were learned, recognition performance was better for views spanned by the training views than for any other views [15]. This is quite surprising since different views of an object are often more different than views of even different object categories. It is not yet clear how the human visual system builds a common representation of a particular object or object category (like the chairs in Figure 1.2). Wallis and Bülthoff conclude that this is possible due to the temporal contiguity of the visual input [118]. When we see a novel object we usually walk around or turn the object in our hand so that the image distance between consecutive views is usually small and the temporal contiguity serves to associate all these views with a single object. Therefore, object recognition in human psychophysics turned away from the structural description approaches [12] toward view-based recognition. These recent findings from human psychophysics, neurophysiology and computational vision provide converging evidence for a view-based recognition framework in which objects and scenes are represented as collections of viewpoint-specific local features rather than 2-d templates or 3-d models [66]. So the recent decade saw a gradual shift away from the 3-d object reconstruction approach pioneered by Marr [59] toward view-based approaches [17]. To deal with the variability of single images, one general approach has been to move away from the single pictorial level and generate more abstract and view-independent representations [116].

The view-based approach relies on the fact that the variability in the set of views belonging to a single object is still governed by regularities that can be captured at the pictorial level. More generally, the shape of the object is described via the view based approach, where the different views, possibly taken with different acquisition systems, form a non-explicit model of the object. The views compromising the representation of a single object are not merely a collection of independent 2-d object views.

1.3 Hierarchical Matching

In order for view-based representations to generalize between exemplars or between views, robust matching algorithms must be specified. In general, the term *robust* describes a system that has demonstrated an ability to recover from the whole range of exceptional inputs and situations in a given environment. A vision procedure is said to be robust if small changes in the assumed model on which the procedure or technique was developed, produce only small changes in the result [28]. Data which do not fit the assumed model and which in fact are very far from fitting the assumed model, constitute a small change in the assumed model. Robustness thus guarantees that the assumed and the actual model

can be matched even if the assumed model deviates from the correct model, or if the model is an idealized approximation to reality.

Thus, there must be some mechanism to measure the perceptual similarity between an object within the input image and the set of known objects [57]. One possibility would be to simply measure the intensity values (local pixel brightness) as similarity between images, but such representations are not robust enough since brightness is highly unstable over image transformations. To have a more robust representation at this step, either local or global shape features are used to allow a robust matching across images. This approach is again deduced from human psychophysics since our perception of an object's shape is closely correlated with its identity. This effect, also called *shape constancy*, is a property of objects that people usually perceive as constant despite changes in viewing perspective. It has been shown that we perceive the same object as having the same shape when observed from different viewpoints [68]. Shape constancy may occur because we are able to recognize the same object from different perspective views by using different features. For example, we recognize a known person's face both from a front view and a profile view of his or her head. In addition, the facial expressions of the face do not influence the recognition. If one recognizes different views and expressions as the same person, its general shape must also be the same (deviations of the assumed model for recognition by facial expressions prove that human recognition is robust, small changes of the shape do not influence recognition). On the other hand it would be much more complicated to solve the same task on a complete stranger's face or of some unknown object.

To recognize unknown objects from different viewing directions, it is necessary to distinguish some particular parts of the shape like "front" and "back" and the presence of axes of symmetry allowing us to perceive the object shape relative to these axes, and allowing the object to be perceived constant despite different viewpoints [109, 30]. Although it is not yet clear how the visual system is able to overcome perspective differences in perceiving object shapes, there is no doubt that it attempts to do so, since otherwise we would see the same object from different viewpoints as completely different and classify them as different objects.

So far we have investigated the view-based recognition process that uses shape features as matching strategy. But to match features, we need a model to which the features can be matched. Most of the view-based models incorporate what is referred to as implicit structural information [111]. The term implicit denotes the fact that this type of structural information does not provide a global description of object shape, but instead simply codes relations between local features (or local shape features). Therefore, a suitable model holds implicit structural information regarding the spatial relations between local features [63]. The form of the structural information is not a global description of an object in term of parts, but is a relatively local description that captures the positional certainty between image measurements [23].

To perform recognition one has to match the actual image with the a priori given or learned model. There exist two different basic strategies for the matching process:

- **Bottom-Up** processing or more descriptively called data-driven processing refers to processes that take a "lower-level" representation as input and create or modify a "higher-level" representation as output. A bottom-up design provides a quick solution for a specific problem by solving it case-by-case, in an ad-hoc way. This solves the specific problem with a diagnostic method: First there is an interactive experimentation stage, during which given datasets are tested with existing tools. In this stage all parameters are evaluated and adjusted in order to solve the given task in the best way and then the program code is adapted based on trial and error. The advantage of this strategy is that the solution is compact and simple, just sufficient to solve the specific problem. The central problem of this quick solution lies in solving a very specific case. For each application all possible techniques and existing tools have to be explored to find the most optimal solution that solves the problem. The analysis process gained in the specific case can rarely be re-used.
- **Top-down** processing, also called hypothesis-driven or expectation driven processing, refers to processes that operate in the opposite direction of bottom-up processing, taking a "higher-level" representation as input and verifying, producing, or modifying a "lower-level" representation as output. A top-down design starts with the definition of the problem space in which the specific problem is embedded. 'Legal' changes in the input data are only specific aspects of the problem space. This can be seen as a development of a general algorithm for a restricted domain, which needs to deal with a variety of possible different problems. The problem space can be described by an abstract language which covers both the possible inputs and the possible outputs. The analysis successively refines the abstraction until operators can be applied to data. As a result such systems can get very complex, since they usually have many different algorithms with a tremendous number of parameters, that need to be tailored to solve each specific problem.

An analogy for these processing techniques can also be found in the human visual system. A naive intuition would state that vision is essentially a bottom-up process. It begins with the sensory input of the retinal images and goes "upward" to perceptual and then conceptual interpretations [68]. However, the perception of the present state of affairs produces expectations about the future and these expectations imply a top-down component to visual processing, because prior high-level interpretations influence current processing at lower level. For example, if one is reading, one expects already "meaningful" words and reads them much faster than meaningless letter- or word strings. As we see, vision (and in this respect recognition) is a mixture between bottom-up and top-down strategies in order to maximize performance.

We can define this mixture in strategies also from the recognition point of view by the kind of information represented in the hierarchical recognition scheme. Doing so we can determine 4 levels or stages of recognition [68]:

- Image-based²,
- surface-based,
- object-based, and
- class-based

levels of perception or recognition.

Figure 1.3 shows the hierarchical recognition scheme together with the top-down and bottom-up approaches. Light is projected onto the retina which generates the retinal image. The retinal image is then processed within the 4 layers and provides the final recognition (bottom-up) in combination with the knowledge incorporated to perform the matching for recognition (top-down). Each of the perception levels contributes to the final recognition based on the retinal image:

Image-based recognition: The output of the first level is called *primal sketch* [58] which is the result of the elementary detection process and includes edges, bars, blobs, and lines and global grouping among the local image features. The structure of such a representation is defined by:

- image-primitives: they represent information about the 2-d structure of the luminance image, such as edges, corners, lines, or other shape primitives,
- 2-d geometry: the geometry of the spatial information among primitives,
- and the reference frame: the coordinate system in which the 2-d shape features are located.

Surface-based recognition: This second level of the recognition scheme is concerned with recovering the intrinsic properties of visible surfaces in the external world, that might have produced the shape features discovered in the image-based stage. Here the 2-d projections are matched within a 3-d coordinate system. Again we have surface-shape-primitives representing the 3-d shape and the 3-d geometry as structure of the representation. Within this stage we have of course only a 2.5-d representation of the primitives.

²Here the term image-based describes the first stage of vision, i.e. the image on the sensor and the image preprocessing operations such as edge detection and linking.

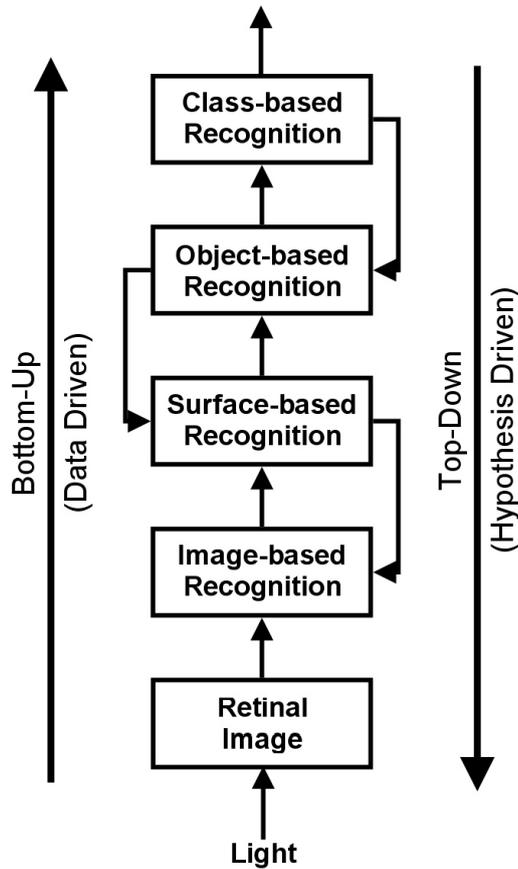


Figure 1.3: Different levels of perception of recognition.

Object-based recognition: In the third level the real 3-d model is matched with the surface-primitives produced in the second stage. Unseen surfaces are included to form a real 3-d object, the structure that holds the information is formed by volumetric-shape-primitives, 3-d geometry and object-based reference frames, which is the coordinate system within which the relations among the shape primitives is given.

Class-based recognition: The highest recognition level is concerned with recovering the functional properties of objects which are accessed through a process of classification. The three previous stages classify an object as being a member of one of a large number of known classes. This identification allows then access to all information about this type of object, including its function and expectations about its future behavior.

As already mentioned, recognition is not organized strictly in a bottom-up manner, but all different levels interact with each other to enhance the result of recognition. If many shape-primitives for instance indicate that the object might be a chair, the knowledge about the class chairs tells us that it might be a chair we can sit on or it might also be an image of a chair on which we cannot sit. To decide this, the information is down-projected and it is checked whether the image-shape features produce a 2-d or 3-d representation of the chair (that is checking the disparity within the 2 retinal images). This example shows that the shape-features within the different levels are extremely important to perform recognition. Furthermore, the hierarchical structure is necessary to separate relevant information and to check hypotheses [91]. All different levels have to adapt to changes in the representation and the visual input to robustly recognize all classes of objects.

We conclude that the desirable features of a shape-based machine vision system are:

- view-based,
- hierarchically structured,
- modular, and
- robust.

This habilitation thesis is a collection of different papers, each describing a vision or recognition task for a specific type of application which fulfill most of the above mentioned properties.

1.4 Summary of Papers

The papers in this collection have been chosen as representative papers for the work conducted in the area of shape based machine vision, where two main areas can be identified: hierarchical recognition and classification and 3-d data acquisition and computation, which have been the main focus of my recent research work. All of them are either book, journal or IEEE proceedings publications. Four papers deal with the level based recognition structure, based on the *General ANalysis Graph (GANAG)*. The remaining two papers describe recent work performed in 3-d vision as part of level 3 (object based recognition). The collection starts with the paper on the GANAG and continues with 3 applications of the GANAG on three real word problems in industrial visual inspection, archaeology, and art history. For visual inspection and archaeology the application of the GANAG is shown in 2-d whereas in art history the applicability of the concept to

3-d is shown. The range image analysis continues with a paper on registration of range data where we proposed a novel technique for range image registration. The collection ends with a work on combining different 3-d acquisition techniques within the hierarchical framework.

All papers were chosen because they represent an overview on the specific research or application area and do not focus on details of the complete framework. The paper on art history, for example, is an invited paper for a mixed machine vision and art history audience, the paper on inspection is written for people working on machine vision applications, and the archaeology application paper for both, the archaeology and machine vision community. Let us now briefly summarize the content of the 6 papers, each denoted by a name to be referred to in the discussion in the following section.

GANAG framework: In the first paper [78] the framework of the General ANALysis Graph (GANAG) is presented. The paper discusses how the hierarchical, shape-based graph is built and used in order to recognize and verify objects. It shows how the *shape* based analysis can be integrated for specific machine vision applications. This paper concentrates on the GANAG within a systematic automated visual inspection concept that speeds up the development of such systems by increasing the flexibility. The detection of primitives is separated from the analysis process. The paper describes a novel strategy for detection, recognition and inspection of objects by introducing the primitives concept where every primitive has a priori known parameters: shape, relative position and size. Based on the shape of the primitives, pattern recognition algorithms are selected to detect the primitives in intensity images. Together with an object-specific shape based description, the analysis graph is instantiated to perform the inspection. The analysis graph can be seen as a "recipe" for solving industrial applications, stating which kind of decisions have to be made at which stage. It is shown that this novel systematic approach also permits a high degree of flexibility since application specific and application independent parts are separated.

Inspection: The paper [87] deals with the problem of application constraints in the design of machine vision systems. In the example of a successful application in the field of visual inspection, the automatic calibration of analog display measuring instruments using the GANAG concept is presented. In this paper the general recognition theory for many different types of measuring instruments is extended and described, considering both industrial constraints and the fact that no redesign of the working process should be necessary if, for example, another measuring instrument is used, or if the pointers have different colors, or if the illumination conditions differ, or if the measuring instrument is rotated or if other changes occur. The focus of the paper lies in the discussion of specific constraints of the application and

the industrial environment in order to refine the general design to an applicable and efficient device by modifying both hardware and software configuration depending on given constraints. The paper shows how the design of the analysis process refines the coarse analysis process by adding constraints generated by the specific application and the industrial environment, which are: speed, cost, accuracy and reliability.

Archaeology: The paper [40] deals with a shape-based classification based on primitives. This time the GANAG is applied to 2-d projections of 3-d objects in the field of Archaeology. 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. This 2-d representation of the former 3-d objects allows the extraction of the general shape properties of the object also from its parts. The complete reconstruction can be performed correctly since the objects are rotationally symmetric, as they have been manufactured on a turntable. In our approach this profile line is segmented into shape-primitives that are again encoded in the analysis graph used for reconstruction. We present a hierarchical segmentation of the profile into rim, wall, and base by providing segmentation rules based on expert knowledge of archaeologists and the curvature of the profile. The profile primitives in this specific application are represented by spline functions, in order to use shape parameters for classification and recognition.

Art history: The extension of the GANAG into the 3-d domain is shown in [84]. In this paper, the hierarchical classification is applied in the field of art history. A hierarchically structured classification scheme is introduced, which separates the classification into three different levels of information: color, shape of region and structure of brush strokes. To compare different regions in painted portraits, it is necessary to project them into a 3-d reference coordinate system, since the faces usually look into different directions and have thus to be normalized into a standard direction via a 3-d face model. The classification scheme allows a mixture between a global top-down classification and a local bottom-up verification in each classification step encoded in the analysis graph. The 3-d head shape information is inevitable to relate different face regions of different portraits to one another.

Registration: In [86] a solution to the problem of 3-d registration of surfaces without corresponding points is presented. The typical 3-d scanner output are range images of objects from one direction at a time. These range images have to be registered to one another in order to reconstruct the complete object in 3-d. The geometric alignment of two three-dimensional surfaces is performed using a modified Iterative Closest Point algorithm, which needs an initial estimate of the relative pose. This initial estimate is computed by taking into account the shape properties of the original object. In this paper we propose a pre-alignment algorithm for registering the front- and back-view of rotationally symmetric objects from range data. For this

kind of object, the surfaces to be registered have to be pre-aligned carefully because otherwise pair-wise registration techniques fail, since there are no corresponding points in the range images. We show a novel model-based technique that computes and uses the axis of rotation of fragments belonging to the same rotationally symmetric object to bring two views of a scene into alignment.

Combination: In [101] a novel technique for combining two different, shape based 3-d acquisition techniques, Shape from Silhouette and Shape from Structured Light is presented. This method solves the problem of acquiring the 3-d shape of objects with handles or similar hole building shape properties. The 3-d reconstruction is based on a sequence of images of the object taken from different viewpoints with shape from silhouette and shape from structured light. The output of both algorithms are then used to construct a single 3-d model. The paper focuses on the 3-d modeling of objects using a voxel based representation. The algorithm proposed employs only simple matrix operations for all the transformations and it is fast since the model generation is faster than the acquisition time. It is shown that the shape of an object cannot be represented exactly using only one acquisition method, since every acquisition method acquires only some properties of the shape whereas others can only be acquired using other acquisition techniques.

The six papers selected represent one specific aspect of shape based machine vision. On the subject of every representative, other papers were also published. They focus on specific aspects of the topic, and cannot be included in this habilitation thesis since they either do not focus on shape or machine vision or describe specific details instead of an overview. Details on the six main topics can be found in:

- GANAG framework: [74, 90, 91, 76, 89, 92, 77, 96].
- Inspection: [88, 73, 75, 81, 79].
- Archaeology: [98, 99, 93, 94, 61, 95, 100, 34, 41, 35, 36, 34, 41, 38, 29, 18, 56, 42, 37, 1, 39, 56, 43, 43].
- Art History: [80, 82, 112, 103, 83, 31].
- Registration: [85, 97, 33, 32].
- Combination: [102, 44, 113, 114, 45, 46, 46, 47, 115].

1.5 Discussion of the Papers

The papers I have selected for the habilitation thesis are related to each other in various aspects. First of all, all papers deal with recognition issues within the 4 different recognition levels. Table 1.1 shows which recognition level is covered in which paper concisely.

Recognition level	GANAG framework	Inspection	Archaeology	Art history	Registration	Combination
Image-based	X	X	X	X		
Surface-based	X	X	X	X	X	
Object-based			X	X	X	X
Class-based	X			X		X

Table 1.1: Levels of recognition described by papers (the “X” marks the level discussed in the paper).

The “GANAG framework” paper shows how all levels of the recognition framework interact with one another and how the different levels can be modeled in order to allow an adaptive representation of the objects to be modeled. However, the results shown use only the first two levels and the top one since only 2-d objects are recognized and modeled.

The “inspection” paper tackles mainly the image-based recognition problem, since it describes the recognition of primitives for a real application and how these primitives are recognized in order to read the value the instrument is displaying. The shape of the image-primitives guides the recognition process, which is limited in this application to the second recognition level since an a priori known boundary condition limits the recognition, which is necessary for machine vision applications.

The “archaeology” paper can be categorized as image-based recognition because shape primitives have to be segmented from one another. A 3-d scanner produces data for this segmentation. They are then reduced to 2-d profiles in order to represent the 3-d data. This can be regarded as back-projection from level 2 and 3 into the first level, where primitives must be found in order to perform classification and therefore recognition. The exact classification is only possible if the data collected in level 1 and 2 is correctly segmented and correctly linked to already existing descriptions. Eliminating one dimension while preserving the 3-d shape ensures a more efficient description and matching, because only necessary information is kept to describe the object.

The “art history” paper uses all 4 recognition levels since the heads and faces are 3-d objects. Furthermore, it shows that the extension of the analysis graph into 3-d is possible.

The “registration” and “combination” papers fit into the area of model generation, which is the main task of level 3, the object-based level. 3-d part data is combined in order to form the complete 3-d description. The “combination” paper shows how different image-based levels and acquisition techniques can be fused in the surface-based level in order to provide more accurate and robust data for level 3, the object-based representation. Here it is shown that full 3-d information for the model generation can be acquired and combined.

Thus, all selected papers fulfill either completely or partly recognition tasks. Let us now discuss their significance within the research community. Two of the papers (inspection and art history) were published in 1994 and 1998 respectively, thus these papers have already influenced other researchers. In the visual inspection research area, most of the work is unpublished since companies prefer that their algorithms remain confidential, so some of the companies used the framework (like ENEL, the Italian power company for example), but have not reported these applications. Recently Correa Alegria and Cruz Serra [2] reported one application for analog and digital display instruments based on the “inspection” and “GANAG framework” paper. In art history, only a few publications deal with the specific area of artist authentication based on brush strokes [55], where our approach is reported by Maitre et. al. (a previous version of the “art history” paper).

The other three papers are quite recent, but there are some previous versions cited by other researchers. For the “archaeology” paper a previous version [100] was cited by researchers like Leymarie et.al. [51], Papaioannou et.al. [69], Cooper et.al. [25], and da Gama Leitao and Stolfi [20], all working in the area of virtually classifying and assembling archaeological fragments. Our approach in this area is significantly different since we mimic the strategy archaeologists use to classify pottery and are thus able to find matching pieces via the classification scheme, whereas others have to match the outline (broken edge) of every fragment with every fragment. As basis for the classification we use the profile line as archaeologists do. Therefore we need the axis of rotation, which is computed via the novel method proposed in the “registration” paper, where earlier work (like [32]) is cited for instance by El-Hakim [24] and Papaioannou et.al. [69].

To our knowledge, the volume-based combination of shape from structured light and shape from silhouette (“combination” paper) is a novel approach in this area. All other approaches to shape from silhouette reported show that the technique is only feasible for convex objects, our extension allows the acquisition of arbitrary objects with this combination.

Let us now discuss how these papers fulfill the properties we have required from recognition systems, i.e.,

- view-based,

- hierarchically structured,
- modular, and
- robust.

View-based: This property is fulfilled by all papers since all of them use single images that are then used either for recognition or to build a model upon. The GANAG is generated using a high number of different views of the same object all devised from different images. We have shown that using these multiple views the representation is robust enough to perform recognition for industrial applications in the “inspection” paper and in [75, 79]. The same strategy has been used in the “art history” paper, here a 3-d model was constructed out of different 2-d images of faces. In the “archaeology” paper we show that multiple views of the same object give the same, consistent representation in the third recognition level, which is the condition needed for subsequent classification [41, 56]. The level 3 papers, “registration” and “combination” inherently use properties of different views in order to generate the model using shape dependent and shape independent techniques to perform this task. We also showed that the number of views taken can be limited using a next-view planing technique [53].

Hierarchically structured: All the methods based on the GANAG concept use the hierarchical structure to perform recognition and classification. The four recognition levels are not present in all of them, however. If applications have a very narrow application field it is not necessary to use all of them, since boundary conditions (like limited number of objects) already provide a priori hypotheses which are normally generated by the fourth level of recognition. So the “inspection”, “archaeology”, “art history” papers and of course the “GANAG framework” paper use the hierarchical model representation and recognition strategy. The “registration” paper also uses hierarchical information since the complete process is model-based. The underlying hypothesis is the rotational symmetry of the objects that can use this technique of registering views without corresponding points in different views. The computation of the axis of rotation can be seen as extracting the global shape information of the surface in order to represent the surface by a single vector and a 2-d function. So the surfaces are first generalized and this generalization is then used to perform registration on the actual data. The hierarchical structure in the “combination” paper is given by the data structure used to describe the final model. The coarse octree model produced in the first level is consequently refined until the final model is derived using both of the input data. Here we have a coarse to fine model generation, where the search and refinement hierarchy is given by the model graph [47].

Modularity: Since all papers are within the level-based recognition strategy, the modularity is inherently given. The intention for the development of the GANAG framework was the adaptivity of the concept in order to guarantee its applicability. This property is reached by separating the detection of primitives from the model-based analysis process. Together with an object-specific description, defined in a so-called description language, the analysis graph is instantiated [96, 100]. Existing software is re-used in the detection stage and therefore **the use of any detection algorithm is possible without changing the analysis**. The modular concept is based on a systematic approach using generic detection algorithms. Again the “registration” and “combination” papers have to be considered separately. The “registration” paper shows the technique as a module itself, which consists of further modules, namely the axis of rotation determination and the ICP-based registration, which can also be used separately. The technique is part of the third recognition level and can be seen as a module that gets view-based data, computes the underlying structure and the registered model as a result [97]. For the “combination” technique we have shown the modularity by using two different acquisition techniques with the same model. The modularity lies in this extendibility since the first attempt was performed with shape from structured light only [52] and was extended to incorporate the silhouette module. Furthermore, an additional technique, the shape from stereo method could be integrated into the same framework [21, 62].

Robustness: The GANAG method has been explicitly designed to improve the robustness of detection and recognition. Robustness is introduced using a library of different detection algorithms, parameter adjustment algorithms and the analysis graph instantiation, the series of test images and evaluating the results. For analogue display instruments the method proved to be robust since the reliability was more than 99%. This means that 99% of the images were automatically and correctly read. The remaining 1% was left for manual inspection due to distortions and occlusions. The manual inspection was possible in 60% of the cases, the rest had to be left unreadable. Thus the robustness required in the design was reached. In the “art history” paper, the robustness of the general analysis graph framework was extended using further features of the object (not only shape), namely color and stroke positions. For the profile classification in the “archaeology” paper, the robustness of the shape (profile) computation is given by the spline based representation of the profile and the subsequent matching criteria that allow tolerances. The robustness of the registration method can be shown by using a robust method for axis determination (up to 15% of the surface normals can be wrong) and the modified ICP algorithm. The combination of shape from silhouette and shape from structured light (“combination” paper) ensures a robust surface shape computation since surface points are measured by two different acquisition techniques and multiple views, so that outliers and wrong surface points can be detected.

The six papers selected for this habilitation thesis show, as representatives for the six main areas of my research work, that all methods proposed fulfill the required properties of shape-based machine vision and recognition methods. They have been chosen because they give a good overview on the research area on the one hand, on the other hand they show that the shape of the object is the key-attribute to perform recognition and classification. Furthermore, the application areas show that the theory given in the GANAG framework and the 3-d data acquisition and registration is applicable to real-world problems like visual inspection, archaeology and art history.

Doing research in the application domain not only gives interesting insights into interdisciplinary working aspects (the vocabulary of scientists of different subject areas like vision and archaeology is significantly different), but also raises interest in both subject areas and influences other vision scientists to work in this specific field. Thus, the research work is growing into a new field, like the cultural heritage field on the example of archaeology, where the work on archaeological fragments has led to an EU project [18] and motivated other researchers to work in the same field (like Leymarie et.al. [51], Papaioannou et.al. [69], Cooper et.al. [25], and da Gama Leitao and Stolfi [20]). This trend can also be made out at major conferences like CVPR and 3DIM, where special workshops on the application of computer vision in archaeology and cultural heritage are introduced (I have the pleasure to be in the program committees).

In the Machine Vision area there is a great demand for new concepts, since in industry vision tasks improve automation. The GANAG approach ensures that costs are reduced due to re-usable systems. At present, complete systems are not available; there are only prototypes in use and almost no new concepts can be found in the literature. The main reason for that is company politics; developments made by so-called Machine Vision Companies are not reported in the literature, since competitors in the market could use these concepts and sell them for a better price. In the academic field only a few researchers are working on systematic AVI, the majority is trying to adapt computer vision algorithms to machine vision applications (see [67]), struggling with system engineering problems [10]. Therefore, work on this research topic helps to advance progress in Machine Vision.

There also exist many publications in the field of shape-based registration techniques, which indicates that it is an important research area and necessary for every scanning software. Users want to have a final “nice looking” 3-d model of the scanned object and do not want to revise these models manually (which is up to now not possible). There are two main classes of approach to fusing multiple 3-d datasets with no clearly superior approach so far. The two main classes are based on fusing surface-based representations (such as triangulation) [110], or fusing multiple point sets into volumes [19], or fusing volumes created from multiple point sets individually [119]. Many approaches do registration first and then fuse the data as a final step, but iterative registration and fusion processes should be investigated in the future [13].

As we see from the registration example, in order to really reach the competence of biological vision and recognition systems a great deal of work remains to be done. The work done so far in the GANAG framework helps to solve some specific tasks, however there are many more to be solved. To be fully acceptable in the inspection area, concepts must be developed that make the implementation of applications easier. For the GANAG framework, the extension to 3-d primitives could open a large field of applications of the concept in the robot vision area. Since the generic detection can handle sparse and dense data, the concept is applicable to grasping and the hand eye problem of robots by defining primitives in range images. Another interesting challenge would be to solve the problem of automatic detection algorithm selection and parameter adjustment using knowledge based decision strategies in the first level of recognition.

Up to now, the GANAG framework handles only still images. For a biologically motivated machine vision system, however, it is necessary to enhance the framework to image streams and motion which is important in the area of navigation and robot control. This extension of the GANAG can also be considered as a movement into the research area of Cognitive Vision, which describes computational processes transforming a video signal into a natural language text describing the spatiotemporal development within a recorded scene [64]. A Cognitive Vision System comprises not only a numerical (shape-based) description of the scene state, but in addition a conceptual description, together with an algorithmic inference engine which allows the manipulation of this conceptual description based on logic operations [64]. Since Cognitive Vision is not necessarily active (only in the learning phase) [26], the GANAG framework can be modified in order to handle also image streams. Therefore, future work will also be directed to the development of the conceptual description.

There are many more properties that need to be considered, e.g., texture, color, temporal aspects, etc. The four proposed levels of recognition represent the current best guess about the overall structure of visual perception. It is conceivable that someday we will attain a full enough understanding of the human visual recognition system, and discover and understand the biological mechanisms to some significant extent, so that these findings can further influence the development of fast and reliable machine vision systems that can work in unconstrained environments.

Bibliography

- [1] 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 Intl. Workshop on Archaeology and Computers*, pages 3–12, Vienna, Austria, 2001.
- [2] F. Correa Alegria and A. Cruz Serra. Automatic calibration of analog and digital measuring instruments using computer vision. *IEEE Trans. on Instrumentation and Measurement*, 49(1):94–99, February 2000.
- [3] Y. Aloimonos. Purposive and qualitative active vision. In *Proc. of 10th International Conference on Pattern Recognition, Atlantic City*, volume I, pages 346–360, 1990.
- [4] Y. Aloimonos. What i have learned: Reply. *Computer Vision Graphics and Image Processing*, 60(1):74–85, 1994.
- [5] Y. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1988.
- [6] J.R. Anderson. The adaptive nature of human categorization. *Psychological Review*, 98:409–429, 1991.
- [7] B.G. Batchelor. *Pattern Recognition - Ideas in Practice*. Plenum Press, New York, London, 1978.
- [8] B.G. Batchelor and D. Braggins. Commercial vision systems. In C. Torras, editor, *Computer Vision: Theory and Industrial Applications*, pages 405–452, New York, 1992. Springer Verlag.
- [9] B.G. Batchelor, D.A. Hill, and D.C. Hodgson. *Automated Visual Inspection*. IFS, Bedford, UK and North-Holland, The Netherlands, 1985.
- [10] B.G. Batchelor and P.F. Whelan. Machine vision systems: Proverbs, principles, prejudices and priorities. In *Proc. of Applied Machine Vision Conference, Cincinnati*, pages 7–19, 1996.

- [11] B. Bhanu. Special issue on innovative applications of computer vision. *Machine Vision and Applications*, 7(3):125–218, 1994.
- [12] I. Biederman. Recognition-by-component: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.
- [13] G. Blais and M.D. Levine. Registering multiview range data to create 3d computer objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(8):820–824, August 1995.
- [14] R.A. Brooks. Intelligence without reason. In *AILAB Memo*, volume 1293. MIT AI Lab, 1991.
- [15] H.H. Bühlhoff and S. Edelman. Psychophysical support for a 2-d view interpolation theory of object recognition. In *Proc. of the National Academy of Science*, volume 89, pages 60–64, 1992.
- [16] H.H. Bühlhoff, S.-W. Lee, T. Poggio, and C. Wallraven. *Biologically Motivated Computer Vision*. Springer, Berlin Heidelberg New York, 2002.
- [17] H.H. Bühlhoff, C. Wallraven, and A. Graf. View-based dynamic object recognition based on human perception. In R. Kasturi, D. Laurendeau, and C. Suen, editors, *Proc. of 16th International Conference on Pattern Recognition, Quebec City*, volume 3, pages 768–776. IEEE Computer Society, 2002.
- [18] J. Cosmas, T. Itagaki, D. Green, E. Grabczewski, L. Van Gool, A. Zalesny, D. Vantintel, F. Leberl, M. Grabner, K. Schindler, K. Karner, M. Gervautz, S. Hynst, M. Waelkens, M. Pollefeys, R. DeGeest, R. Sablatnig, and M. Kampel. 3D MURALE: A multimedia system for archaeology. In *Proc. of Intl. EuroConference on Virtual Reality, Archaeology and Cultural Heritage*, page in press, Athens, Greece, Nov. 28–30 2001.
- [19] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proc. of SIGGRAPH, New York*, pages 303–312, 1996.
- [20] H.C. da Gama Leitao 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, September 2002.
- [21] M. Dangl and R. Sablatnig. Trinocular stereo for robotic arm control. In M. 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 225–233. Oldenbourg Wien, München, 1999.

- [22] N. Duta, A.K. Jain, and M.P. Dubuisson-Jolly. Automatic construction of 2d shape models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(5):433–446, May 2001.
- [23] S. Edelman and D. Weinshall. A self-organizing multiple-view representation of 3d objects. *Biological Cybernetics*, 64:209–219, 1991.
- [24] S. El-Hakim. A practical approach to creating precise and detailed 3d models from single and multiple views. In *International Archives of Photogrammetry and Remote Sensing*, volume 33, pages 122–129, 2000.
- [25] D. B. Cooper et al. Bayesian virtual pot-assembly from fragments as problems in perceptual-grouping and geometric-learning. In R. Kasturi, D. Laurendeau, and C. Suen, editors, *Proc. of 16th International Conference on Pattern Recognition, Quebec City*, volume 3, pages 297–302. IEEE Computer Society, 2002.
- [26] G.H. Granlund. Does vision inevitably have to be active? In *Proc. of Scandinavian Conference on Image Analysis*, volume I, pages 138–146, 1999.
- [27] A. Gray. The Möbius Strip. In *Modern Differential Geometry of Curves and Surfaces with Mathematica, 2nd Edition*, pages 325–326, Boca Raton, FL, 1997. CRC Press.
- [28] R.M. Haralick and G.L. Shapiro. Glossary of computer vision terms. *Pattern Recognition*, 24(1):69–93, 1991.
- [29] K. Hlavackova-Schindler, M. Kampel, and R. Sablatnig. Fitting of a closes planar curve representing a profile of an archaeological fragment. In *Proc. of Int. Euro-Conference on Virtual Reality, Archaeology and Cultural Heritage*, page in press, Athens, Greece, November 2001.
- [30] D.P. Huttenlocher and L.M. Lorigo. Recognizing three dimensional objects by comparing two-dimensional images. In *Proc. of Intl. Conf. on Computer Vision and Pattern Recognition*, pages 878–884, San Francisco, CA, 1996.
- [31] P. Kammerer and R. Sablatnig. Head pose estimation in painted portraits used for comparison. 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 127–134, Oldenburg, Wien, München, 2001.
- [32] 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, Muenchen, 1999.

- [33] M. Kampel and R. Sablatnig. Range image registration of rotationally symmetric objects. In N. Braendle, editor, *Computer Vision - CVWW'99, Proc. of the Computer Vision Winter Workshop, Rastendorf*, pages 69–77, Vienna, Austria, 1999. PRIP, TU Wien.
- [34] M. Kampel and R. Sablatnig. Color calibration for pre-classification of pottery. In T. Svoboda, editor, *Proceedings of the Czech Pattern Recognition Workshop 2000*, pages 185–189, Praha, Czech Republic, 2000. Czech Pattern Recognition Society.
- [35] 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.
- [36] M. Kampel and R. Sablatnig. Color classification of ceramics using spectral reflectance. In *Proc. of 1st Intl. Conference on Color in Graphics and Image Processing, Saint-Etienne*, pages 323–327, 2000.
- [37] M. Kampel and R. Sablatnig. Automated 3D recording of archaeological pottery. In D. Bearman and F. Garzotto, editors, *Proc. of the Intl. Conf. on Cultural Heritage and Technologies in the Third Millennium*, volume 1, pages 169–182, Milan, Italy, Sept. 2001.
- [38] M. Kampel and R. Sablatnig. Dreidimensionale Oberflächenrekonstruktion archäologischer Scherben zur Klassifizierung. In G. Stanke and L. Paul., editors, *Proc. of 3D-Nordost 2001: 4. Anwendungsbezogener Workshop zu Erfassung, Verarbeitung, Modellierung und Auswertung von 3D-Daten*, pages 75–82, Berlin, Germany, 2001.
- [39] 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.
- [40] 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.
- [41] M. Kampel and R. Sablatnig. Computer aided classification of ceramics. In F. Niccolucci, editor, *Virtual Archaeology Proceedings of the VAST2000 Euroconference, Arezzo*, pages 77–82, Oxford, 2002. Archaeopress.
- [42] 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.
- [43] M. Kampel, R. Sablatnig, and H. Mara. Automated archivation 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.
- [44] M. Kampel, R. Sablatnig, and S. Tosovic. 3d shape modeling of objects with concavities. In *Proc. of Intl. European Conference on Visual Perception*, page in press, Kusadasi, Turkey, August 26-30 2001.
- [45] M. Kampel, R. Sablatnig, and S. Tosovic. Fusion of surface and volume data. 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 21–28, 2002.
- [46] M. Kampel, R. Sablatnig, and S. Tosovic. Volume based reconstruction of archaeological artifacts. In W. Boehler, editor, *Proc. of Intl. Workshop on Scanning for Cultural Heritage Recording*, pages 76–83, 2002.
- [47] 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 Intl. Symposium on 3D Data Processing Visualization and Transmission, Padova*, pages 754–757, 2002.
- [48] G. Kanizsa. Organization in vision: Essays on Gestalt Perception. In *Praeger*, 1979.
- [49] R. Klette, K. Schlüns, and A. Koschan. *Computer Vision: Three Dimensional Data from Images*. Springer, Singapore, 1998.
- [50] K. Koffka. Principles of Gestalt Psychology. In *Harcourt, Brace and World*, 1935.
- [51] F. Leymarie, D. Cooper, M. Sharp Joukowsky, B. Kimia, D. Laidlaw, D. Mumford, and E. Vote. The shape lab. - new technology and software for archaeologists. In Oxford UK BAR International Series 931, Archaeopress, editor, *Computing Archaeology for Understanding the Past (CAA 2000)*, pages 79–89, 2000.
- [52] C. Liska and R. Sablatnig. Adaptive 3D acquisition using laser light. In T. Svoboda, editor, *Proceedings of the Czech Pattern Recognition Workshop 2000*, pages 111–116, Praha, Czech Republic, 2000. Czech Pattern Recognition Society.
- [53] C. Liska and R. Sablatnig. Estimating the next sensor position based on surface characteristics. 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 1, pages 538–541. IEEE Computer Society, 2000.
- [54] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, 1998.
- [55] H. Maitre, f. Schmitt, and J. Crettez. High quality imaging in museums from theory to practice. In *Proc. of Very High Resolution and Quality Imaging II*, volume 3025, pages 30–39. IS&T/SPIE, 1997.
- [56] 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.
- [57] F. Marques, M. Pardis, and R. Morros. Object matching based on partition information. In *Proc. of 9th Intl. Conference on Image Processing*, volume II, pages 829–832, Rochester, NY, 2002.
- [58] D. Marr. *Vision*. Freeman, New York, 1982.
- [59] D. Marr and H.K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proc. of the Royal society of London*, B-200:269–294, 1978.
- [60] A.D. Marshall and R.R. Martin. *Computer Vision, Models and Inspection*. World Scientific, 1992.
- [61] C. Menard and R. Sablatnig. On finding archaeological fragment assemblies using a bottom-up design. In W. Burger and M. Burge, editors, *Proc. of the 21st Workshop of the Austrian Association for Pattern Recognition (ÖAGM), Hallstatt*, pages 203–208. Oldenbourg Wien, München, 1997.
- [62] C. Menard and R. Sablatnig. Adaptive area-based stereo matching. In *Proc. of the IS&T/SPIE Symposium on Three-Dimensional Image Capture and Applications*, volume SPIE-Vol.3313, pages 14–24, Jan. 1998.
- [63] T. Messer. Model-based synthesis of vision routines. In C. Archibald and E. Petriu, editors, *Advances in Machine Vision: Strategies and Applications*, pages 79–97. World Scientific, 1992.
- [64] H.H. Nagel. Image sequence evaluation: 30 years and still going strong. In A. Sanfeliu, J.J. Villanueva, M. Vanrell, R. Alquezar, J.O. Eklundh, and Y. Aloimonos,

- editors, *Proc. of 15th International Conference on Pattern Recognition, Barcelona*, volume I. IEEE Computer Society, 2000.
- [65] V.S. Nalwa. *A Guided Tour of Computer Vision*. Addison-Wesley Publishing Company, Bonn, Muenchen, New York, Sidney, Tokyo, 1993.
- [66] F.N. Newell and H.H. Bülthoff. Categorical perception of familiar objects. *Cognition*, 85:113–143, 2002.
- [67] T.S. Newman and A.K. Jain. A survey of automated visual inspection. *Computer Vision, Graphics, Image Processing*, 61(2):231–262, 1995.
- [68] Stephen E. Palmer. *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA, 1999.
- [69] G. Papaioannou, E.A. Karabassi, and T. Theoharis. Virtual archaeologist: Assembling the past. *IEEE Computer Graphics*, 21(2):53–59, March-April 2001.
- [70] M. Pavel. An invitation to shape theory. *Pattern Recognition Letters*, 1(1):31–35, 1982.
- [71] S. Pinker. Visual cognition: An introduction. *Cognition*, 18:1–63, 1984.
- [72] M.I. Posner, editor. *Foundations of cognitive science*. MIT Press, Cambridge, MA, 1990.
- [73] R. Sablatnig. Automatische Ablesung von Wasserzählern zur Qualitätssicherung bei der Eichung. In G. Sagerer, S. Posch, and F. Kummert, editors, *Künstliche Intelligenz und Mustererkennung, 17th DAGM Symposium Bielefeld*, volume 6, pages 335–342. Springer Verlag Informatik Xpress, 1995.
- [74] R. Sablatnig. A highly adaptable visual inspection concept by separating detection and analysis process. In *Proc. 1st International Workshop on Knowledge-Based Systems for the (Re)Use of Program Libraries, Sophia Antipolis*, volume 2, pages 25–27. Sophia Antipolis, 1995.
- [75] R. Sablatnig. Visual inspection of watermeters used for automatic calibration. In R.T. Chin, H.H.S. Ip, A.C. Naiman, and T-C. Pong, editors, *Image Analysis Applications and Computer Graphics*, volume 1024 of *Lecture Notes in Computer Science*, pages 518–520. Springer Verlag, 1995.
- [76] R. Sablatnig. Flexible automatic visual inspection based on the separation of detection and analysis. In *Proc. of 13th International Conference on Pattern Recognition, Vienna*, volume III, pages 944–948. IEEE Computer Society Press, 1996.

- [77] R. Sablatnig. A flexible concept for automatic visual inspection. In T. Pajdla, editor, *Czech Pattern Recognition Workshop 1997, Milovy*, pages 87–96. Czech Pattern Recognition Society, 1997.
- [78] R. Sablatnig. Increasing flexibility for automatic visual inspection: The general analysis graph. *Machine Vision and Applications*, 12:158–169, 2000.
- [79] R. Sablatnig. Automatic visual inspection of watermeters using a flexible inspection concept. *Computers and their Applications*, 8(3):170–182, 2001.
- [80] R. Sablatnig and Horst Bischof. Strukturelle Beschreibung von kunstgeschichtlichen Portraitminiaturen. volume 5 of *Informatik Xpress*. Springer Verlag Berlin, 1994.
- [81] R. Sablatnig and C. Hansen. Automatic calibration of watermeters. In Walter G. Kropatsch Franc Solina, editor, *Visual Modules: Proc. of the 19th ÖAGM and 1st SDRV Workshop*, volume 81 of *OCG Schriftenreihe*, pages 231–239. Oldenburg, Wien, München, 1995.
- [82] R. Sablatnig and P. Kammerer. Pinselstrichsegmentation als Basis für eine Klassifikation von Gemälden. In B. Jähne, P. Geissler, H. Hausecker, and F. Hering, editors, *Mustererkennung 1996, 18. DAGM Symposium Heidelberg*, Informatik aktuell, pages 392–399. Springer Verlag, 1996.
- [83] R. Sablatnig, P. Kammerer, and E. Zolda. Hierarchical classification of painted portraits using face- and brush stroke models. In Anil K. Jain, Svetha Venkatesh, and Brian C. Lovell, editors, *Proc. of 14th Int’l Conference on Pattern Recognition, Brisbane*, volume I, pages 172–174, 1998.
- [84] R. Sablatnig, P. Kammerer, and E. Zolda. Structural analysis of paintings based on brush strokes. In W. C. McCrone and R. J. Weiss, editors, *Fakebusters Scientific Detection of Fakery in Art*, pages 222–244. Hansen, Stoughton, Massachusetts, 1999.
- [85] 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*, pages 339–346. Springer, 1999.
- [86] R. Sablatnig and M. Kampel. Model-based registration of front- and backviews. *Computer Vision and Image Understanding*, 87(1):90–103, 2002.
- [87] R. Sablatnig and W.G. Kropatsch. Application constraints in the design of an automatic reading device for analog display instruments. In *Proc. of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota*, pages 205–212. IEEE-Computer Society Press, 1994.

- [88] R. Sablatnig and W.G. Kropatsch. Automatic reading of analog display instruments. In *Proc. of the 12th IAPR International Conference on Pattern Recognition, Jerusalem*, volume 1, pages 794–797. IEEE-Computer Society Press, 1994.
- [89] R. Sablatnig and A. Leonardis. A general approach to automatic visual inspection. In G. K. Lee, editor, *Proc. of the ISCA 9th International Conference on Computer Applications in Industry and Engineering*, pages 56–59, Orlando, Florida, 1996.
- [90] R. Sablatnig and A. Leonardis. A modular automatic visual inspection concept based on generic detection. In *Proc. of Applied Machine Vision Conference, Cincinnati*, pages 339–358, 1996.
- [91] R. Sablatnig and A. Leonardis. New concepts in automatic visual inspection. In N. Pavesic, H. Niemann, S. Kovacic, and F. Mihelic, editors, *Proc. of the 3rd Slovenian-German Workshop and 2nd SDRV Workshop on Speech and Image Understanding, Ljubljana*, pages 321–330. IEEE Slovenia Section, 1996.
- [92] R. Sablatnig, A. Leonardis, and W.G. Kropatsch. A highly adaptable concept for visual inspection. In M. Jamshidi, J. Yuh, and P. Dauchez, editors, *Proc. of the 1st International Symposium on Intelligent Automation and Control ISIAC'96—Recent Trends in Development and Applications*, volume 4, pages 651–656, Montpellier, May 1996. TSI Press, Albuquerque.
- [93] R. Sablatnig and C. Menard. Computer based acquisition of archaeological finds: The first step towards automatic classification. In P. Moscati and S. Mariotti, editors, *Proc. of 3rd International Symposium on Computing in Archaeology, Rome*, volume 1, pages 429–446, 1996.
- [94] R. Sablatnig and C. Menard. 3d reconstruction of archaeological pottery using profile primitives. In N. Sarris and M.G. Strintzis, editors, *Proc. of Intl. Workshop on Synthetic-Natural Hybrid Coding and Three-Dimensional Imaging*, pages 93–96, Rhodes, Greece, September 1997.
- [95] R. Sablatnig and C. Menard. Classification and 3d- reconstruction of rotational symmetric pottery. In Aleš Leonardis and Franc Solina, editors, *Computer Vision—CVWW'98, Proc. of the Computer Vision Winter Workshop*, pages 186–194, Gozd Martuljek, Slovenia, February 1998. IEEE Slovenia Section.
- [96] R. Sablatnig and C. Menard. On representing the object model for automatic visual inspection using a description language. In *Proc. of the IS&T/SPIE Symposium on Machine Vision Applications in Industrial Inspection VI*, volume SPIE-Vol.3306, pages 131–140, San Jose, USA, January 1998.

- [97] R. Sablatnig and C. Menard. On estimating the position of fragments on rotational symmetric pottery. In *Proc. of 2nd Intl. Conf. on 3D Digital Imaging and Modeling, Ottawa*, pages 455–462, 1999.
- [98] R. Sablatnig, C. Menard, and P. Dintsis. A preliminary study on methods for a pictorial acquisition of archaeological finds. In P. M. Fischer, editor, *Archaeology and Natural Science (ANS)*, volume 1, pages 143–151. P. Astroems Foerlag, Gothenburg, 1993.
- [99] R. Sablatnig, C. Menard, and P. Dintsis. Bildhafte, dreidimensionale Erfassung von archäologischen Fundgegenständen als Grundlage für die automatisierte Klassifikation. In O. Stoll, editor, *Computer & Antike*, volume 3, pages 59–84. Scripta Mercaturae Verlag, 1994.
- [100] R. Sablatnig, C. Menard, and Walter G. Kropatsch. Classification of archaeological fragments using a description language. In S. Theodoridis, I. Pitas, A. Stouraitis, and N. Kalouptsidis, editors, *Signal Processing IX, Theories and Applications*, volume II, pages 1097–1100, Rhodes, Greece, 1998. EURASIP.
- [101] R. Sablatnig, S. Tosovic, and M. Kampel. Combining shape from silhouette and shape from structured light for volume estimation of archaeological vessels. In R. Kasturi, D. Laurendeau, and C. Suen, editors, *Proc. of 16th International Conference on Pattern Recognition, Quebec City*, volume 1, pages 364–367. IEEE Computer Society, 2002.
- [102] 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 Archeology*, volume 1049 of *BAR International Series*, pages 93–99. Archeopress, 2002.
- [103] R. Sablatnig and E. Zolda. Rechnergestützte Präklassifikation von Portraitminiaturen. In J. Hensley and G. Stanke, editors, *Proc. of EVA96: Elektronische Bildverarbeitung & Kunst, Kultur, Historie*, volume V, pages 15–16, Berlin, Germany, November 1996.
- [104] P.G. Schyns. Diagnostic recognition: Task constraints, object information, and their interactions. In M.J. Tarr and H.H. Bülhoff, editors, *Object Recognition in Man, Monkey, and Machine*, pages 147–179, Cambridge, Massachusetts, London, England, 1999. MIT Press.
- [105] P.G. Schyns, R.L. Goldstone, and J.P. Thibaut. The development of features in object concepts. *Brain and Behavioral Sciences*, 21:17–54, 1998.

- [106] A. Sekuler. Local and global minima in visual completion: Effects of symmetry and orientation. *Perception*, 23:529–545, 1994.
- [107] R. Sekuler and R. Blake. *Perception*. McGraw-Hill, New York, 4 edition, 2001.
- [108] A. Selinger and R.C. Nelson. Appearance-based object recognition using multiple views. In *Proc. of Intl. Conf. on Computer Vision and Pattern Recognition*, volume I, pages 905–911, Kauai, HI, 2001.
- [109] A. Slater and V. Morison. Shape constancy and slant perception at birth. *Perception*, 12:707–718, 1985.
- [110] M. Soucy and D. Laurendeau. A general surface approach to the integration of a set of range views. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(4):344–358, April 1995.
- [111] M.J. Tarr and H.H. Bülthoff. *Object Recognition in Man, Monkey, and Machine*. MIT Press, Cambridge, MA, London, England, 1999.
- [112] I. Tastl, R. Sablatnig, and W.G. Kropatsch. Model-based classification of painted portraits. In Axel Pinz, editor, *Pattern Recognition 1996: Proc. of 20th ÖAGM Workshop*, volume 90 of *OCG Schriftenreihe*, pages 237–250. Oldenbourg Wien, München, 1996.
- [113] S. Tosovic and R. Sablatnig. 3D modeling of archaeological vessels using shape from silhouette. In *Proc. of the 3rd International Conference on 3-D Digital Imaging and Modeling, Quebec City*, pages 51–58, May 2001.
- [114] S. Tosovic and R. Sablatnig. Volume estimation for objects with concavities. In B. Likar, editor, *Proc. of the 6th Computer Vision Winter Workshop 2001*, pages 49–59, Bled, Slovenia, February 2001. Slovenian Pattern Recognition Society.
- [115] S. Tosovic, R. Sablatnig, and M. Kampel. On combining shape from silhouette and shape from structured light. In H. Wildenauer and W.G. Kropatsch, editors, *Proc. of 7th Computer Vision Winter Workshop, Bad Aussee*, pages 108–118. PRIP, TU-Wien, Feb. 2002.
- [116] S. Ullman. Three-dimensional object recognition based on combination of views. In M.J. Tarr and H.H. Bülthoff, editors, *Object Recognition in Man, Monkey, and Machine*, pages 21–44, Cambridge, Massachusetts, London, England, 1999. MIT Press.
- [117] D. Vernon. *Machine Vision*. Prentice-Hall, Englewood Cliffs, New Jersey, 1991.

- [118] G.M. Wallis and H.H. Bülthoff. Effect of temporal association of recognition memory. In *Proc. of the National Academy of Science*, volume 98, pages 4800–4804, 1998.
- [119] M.D. Wheeler, Y. Sato, and K. Ikeuchi. Consensus surfaces for modeling 3d objects from multiple range images. In *Proc. of 6th Intl. Conf. on Computer Vision*, pages 917–924, 1997.

Chapter 2

Increasing Flexibility for Automatic Visual Inspection: The General Analysis Graph (“GANAG Framework”)

Robert Sablatnig. Increasing Flexibility for Automatic Visual Inspection: The General Analysis Graph. *Machine Vision and Applications*, 12:158–169, 2000.

Increasing flexibility for automatic visual inspection: the general analysis graph

Robert Sablatnig

Pattern Recognition and Image Processing Group, Institute of Computer Aided Automation, Vienna University of Technology, Favoritenstrasse 9/183/2, 1040 Vienna, Austria; e-mail: sab@rip.tuwien.ac.at, <http://www.rip.tuwien.ac.at>

Abstract. The continuing development of machine vision is initiating a change from human to machine vision for inspection purposes. This paper concentrates on a general analysis graph within a systematic automated visual inspection concept that speeds up the development of such systems by increasing the flexibility. The detection of primitives is separated from the model-based analysis process. Together with an object-specific description, the analysis graph is instantiated to perform the inspection. The analysis graph can be seen as a “recipe” for solving industrial applications, stating which kind of decisions have to be made at which stage.

Key words: Automatic visual inspection – Inspection system generation – Directed graphs – Analogue display instruments

set-up by experienced operators, hard- and software development costs, labor, and maintenance costs. The solution to cost reduction is to increase the flexibility of AVI systems, which makes it possible to amortize the development costs by a high number of installed units for different applications. The AVI system design cost has already been reduced by the development of libraries of image-processing algorithms (e.g. Matrox Image Library [21]) or interactive image-processing systems (like Khoros and Cantata [19,34], KBVision [1], and Matrox Inspector [20]), which allow rapid prototyping and the re-use of algorithms used within these systems. However, it is unlikely that the common user of an inspection system will have the relevant image-processing expertise to be able to set up an inspection system by himself. If image-processing systems are to be adopted and used for inspection of a variety of applications, it is essential to reduce the expertise required in the configuration of the inspection system [3].

This paper shows that the use of a general analysis graph that separates the detection from the analysis allows a flexible adaptation of the inspection system discussed in Sect. 2. The inspection model is represented in a description language, shown in Sect. 3. Following the modeling of the application-dependent analysis in Sect. 4 the application of the general analysis graph for the inspection of analogue display instruments is shown exemplary (Sect. 5). The use of the analysis graph for analogue display instrument in the application of calibrating watermeters shows that the general analysis graph can be successfully used in industrial applications (Sect. 6). The paper concludes with results and a discussion of the flexibility of the analysis graph.

1 Introduction

It has now been well over 30 years since several individuals and groups made concerted efforts to automate visual perception in the research discipline of machine vision [24]. From the beginnings researchers tried to convert the results achieved in basic research into applications to prove that their algorithms work. Historically, the first industrial application area was in the manufacturing industry, because there was a strong desire to automate the production process and to control the final product. With an emerging requirement for improved quality control within the manufacturing industry, the use of visual inspection of the manufactured product becomes a necessity, especially to fulfill the ISO 9000 industrial quality standard [18,27]. While visual inspection is high in potential, at present the design and implementation of automatic visual inspection (AVI) systems is labor intensive. In addition, most of the visual inspection systems have been developed in isolation with no systematic approach, which has led to the design of inflexible customized solutions involving very high system engineering costs over the last 20 years [4,5,25].

The major drawbacks of existing inspection systems are high set-up costs, resulting from extensive pre-inspection

2 General inspection concept

One solution in designing a flexible visual inspection system lies in the separation of the application-independent feature detection from the application-dependent analysis, forming the model-based AVI system [29]. The concept of the separation has to be integrated into a complete inspection system that allows an interactive set-up by the user. This interaction is valuable since the specific application knowledge can only be provided by the user, who is familiar with the specific in-

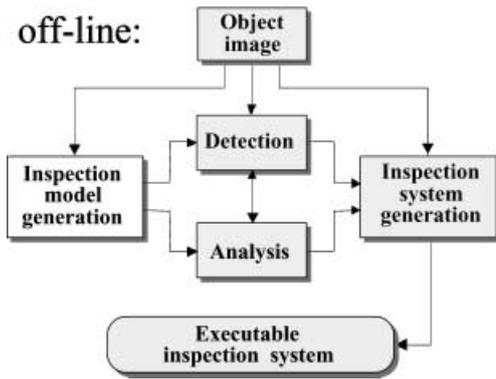


Fig. 1. Inspection system concept (off-line)

spection problem and the inspection hardware available. Figure 1 shows the off-line part of the proposed general inspection system concept (rectangles indicate processes, rounded rectangles data). This part is called off-line since it is not performed at the speed of the production line, there is no explicit time limit to complete the inspection system set-up (except economical set-up time limits). It is supposed that the system set-up takes place at the final working site with the illumination and hardware that should be used in the final system.

The illumination and imaging geometry should be worked out carefully to increase the quality of the source data, influencing the quality of the result. To provide useful source data for the inspection process, the image acquisition properties should not be changed after the set-up. A particularly common error is the tendency to concentrate on the image processing to the detriment of the image acquisition (i.e., pose of the object being inspected, lighting, optics and sensor) [2]. A well-designed sensing system reduces noise, prevents blur, stops object motion, optimizes the contrast between the part and the background, for instance, has a resolution that ensures defect detection in the desired size and emphasizes all features relevant for inspection. Since the success of an inspection system is critically dependent upon the correct engineering of the front end component illumination, presentation device and image capture hardware [10, 23], we assume the image acquisition step as carefully designed.

The generated executable inspection system is used on-line, i.e., it works within the production line flow. The off-line inspection system set-up consists of four general processes.

- *Inspection model generation.* Generic algorithms are used to detect inspection features (the so-called primitives) and (possibly together with an a priori model) the inspection model of the object (the so-called description) is supplied.
- *Detection.* The detection algorithms (like edge detectors, edge-linking algorithms, segmentation algorithms) to detect primitives are selected in this stage. Existing, standard detection algorithms like line detection, circle detection, or ellipse detection can be used, which means that no specific detection software (often also called image-processing software) has to be developed for a specific problem.

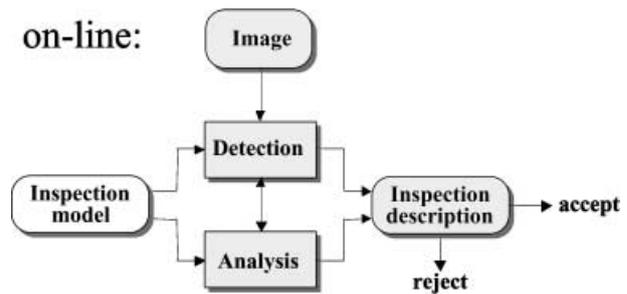


Fig. 2. On-line inspection system components

- *Analysis.* The parameters of the detection algorithms are adjusted, detection order and search space are selected in the analysis stage. Several algorithms are adapted, tested for the given inspection task, and used if they achieve a certain pre-defined detection rate. This step can be seen as the pattern recognition step since it analyses the detected primitives in order to recognize patterns (i.e., the inspection model) defined in the description language.
- *Inspection system generation.* The preliminary inspection system is tested. If it turns out that, for instance, a specific feature detection algorithm does not adhere to the desired recognition rate in the test set because of lighting conditions or alignment, it can be replaced by another algorithm which attains the recognition rate. Furthermore, industrial constraints are checked and balanced before the final inspection system test takes place.

Since the analysis is separated from the detection, the concept allows a flexible adaptation of the inspection system too, because a change of the object layout results only in the requirement of a new description, while the detection and the analysis remains the same. However, if substantial layout changes occur, analysis and detection have to be re-adjusted. If the test confirms that the inspection system attains the intrinsic constraints of the inspection task, the executable inspection system for on-line inspection is ready.

Figure 2 shows the components of the on-line inspection system. The image taken under the same acquisition conditions as in the off-line set-up phase, is the input for the detection. The defined feature detection algorithms are used in the detection order given by the analysis and the generic parameters provided by the description. In the analysis, the features are checked whether they are within the tolerances defined in the description. The inspection result is provided, which states why the object is accepted or rejected.

3 Description language

The object structure (shape primitives and properties) has to be represented in a description language consisting of a graph structure in which nodes represent the primitives and arcs the relations between primitives [28]. A priori information concerning the quality standard (e.g., manufacturing tolerances and detection tolerances) are also part of the model. From the description language point of view, the modeling can be interpreted as a syntactic pattern recognition approach in which the primitives are transformed into the vocabulary and the relations are transformed into a grammar [15].

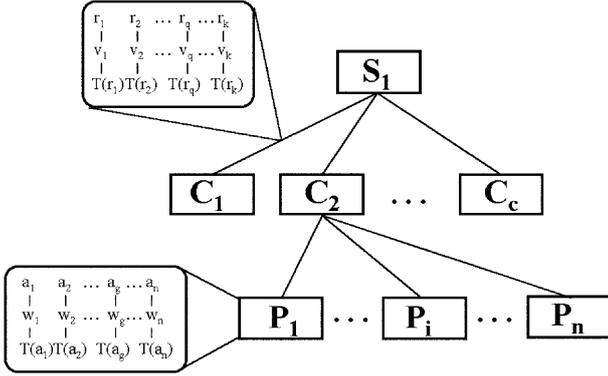


Fig. 3. Description language graph

The proposed approach makes use of the idea of shape decomposition, it divides complex shapes into simple elementary units, i.e., primitives. This concept can be seen as an application of semantic networks [12], since semantic networks are labeled, directed graphs where nodes represent objects, sub-objects, or shape primitives and arcs represent relations between them. A set of attributes that describe different object features is attached to each node; a set of attributes that describe different properties is attached to each arc. Once the object is transformed to this representation, all operations for recognition, verification, and inspection can be executed on this graph structure. The advantage of a description language lies in the uniqueness of representation, different objects result in different descriptions.

Formally, the description language is a graph $G = \langle O, R \rangle$, where $O = \{m \mid 1 \leq m \leq n\}$ denotes the set of nodes and $R = \{\langle c, d \rangle \mid c, d \in O\}$ the set of arcs. A node O consists of different sub-objects or primitives. Each node has different attributes a , with weights w , and a tolerance $T(a)$ defined as

$$T(a) = \begin{cases} 1, & \text{if } |a \bmod -a \text{ img}| \leq c, \\ \frac{1}{|a \bmod -a \text{ img}|}, & \text{otherwise} \end{cases} \quad (1)$$

where c is the allowed tolerance, a^{mod} denotes the value of attribute a in the model, and a^{img} the value of the attribute a in the image

Two nodes are in relation according to R . Each relation $\langle c, d \rangle$ is decomposed into k sub-relations between the same nodes, each with a weight v and a tolerance $T(r)$ defined analogue as for weights and tolerances of attributes. Figure 3 shows the graph and the inner structure of nodes and arcs. Note that all attributes and relations contain numerical values.

The weights w and v are necessary for the model verification. Each of the geometrical, positional, and relational properties has a certain weight in order to verify the corresponding description to a given image. Since these weights are influenced by the data and therefore application dependent, they have to be fixed during the set-up procedure. The verification of image to description consists of verifying whether the number and type of features and primitives are the same. Next, attributes and relations are checked whether they match within given tolerances. The verification process is carried out by comparing all attributes of a node and its

successors with the model. The confidence for a node can be computed based on the result of the comparison:

$$\text{conf}(p) = \sum_{g=1}^n w_g * T(a_g) + \sum_{\langle p,q \rangle \in \mathbf{R}} v_{\langle p,q \rangle} * \text{conf}(q), \quad (2)$$

where w_g are the weights of the attributes of the nodes and $v_{\langle p,q \rangle}$ the weights of the sub-relations of the arcs. Observe that n , the number of attribute values, and m , the number of arcs, depend on the node p . Moreover, for leaves, we have

$$\text{conf}(p) = \sum_{g=1}^n w_g * T(a_g). \quad (3)$$

This 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. The use of weights also allows a two-step identification; primitives or sub-objects with high weights are first detected and checked, next primitives with low weights are postulated on a certain position and then verified.

The hierarchical representation of the object within the description (i.e., shape decomposition) allows an easy update by adding or removing new nodes and arcs since only the node of the subgraph where nodes are added or removed has to be modified the rest of the description remains unchanged. Therefore, small changes in the object layout result only in small changes in the description too.

4 The general analysis graph

In contrast to detection, where independent algorithms detect primitives, the analysis deals with the application-specific knowledge: the description, tolerances and the order of the detection steps. After the generation of the inspection model (i.e., the description), the weights for each primitive of the object, which is a feature of the object within the image, are not fixed, since at this stage it is not clear which primitive can be found reliably and which not, and which primitive is important for recognition. Therefore, one goal of the analysis is the determination of the weights of primitives. To compute weights, the detection order and the parameters of the detection algorithms are evaluated, features with high detection confidence get higher weights than features with low confidence. To model the detection order in the analysis, we use an analysis graph.

Generally, the inspection could take place based on the description. Since performance and speed are crucial and the verifying phase is a graph isomorphism problem whose general case is known to have UP complexity [14], the semantic information stored in nodes and arcs decreases complexity. There are two different strategies for relating features in the image coordinate system to the object-centered coordinate system.

- *Parallel (M1)*. Independent detection and localization of features followed by a verification of the constraints.

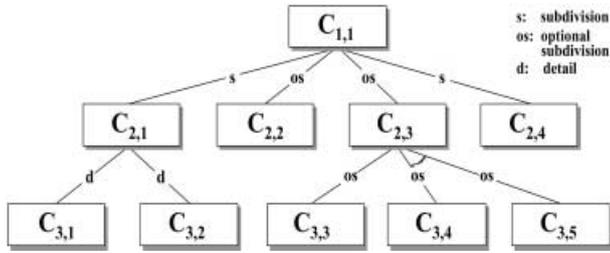


Fig. 4. Analysis graph with three relation types

- *Sequential (M2)*. Detection and localization of reliable features; construction of hypotheses based on imposed constraints of the detected features; verification of hypotheses.

Since each method has its benefits, both principles are used in the proposed analysis process, formulated in a graph structure introducing a hierarchy (see Fig. 4). It represents the space of all possible solutions for the problem in the particular domain. A solution is formulated by instantiating the graph to form a unique solution. Each node represents an element of the solution called *cell*, i.e., an image-processing task like Sobel edge detection or a working step like image acquisition. The *arcs* between the nodes represent one of the following semantic relations (see Fig. 4).

- *Sub-divide relation*. This is an n-ary relation representing a sub-division of a cell into its constituent set of sub-cells. For example, image acquisition can be sub-divided into CCD size, upper and lower threshold for transistor response, digitizer, illumination adaption, and frame transfer to name a few. Each sub-divide relation has a weight, which allows an ordering of relations within the graph, sub-divide relations with equal weights may be executed in parallel. For the example in Fig. 4, the cell $C_{1,1}$ is first sub-divided into $C_{2,1}$ and afterwards into $C_{2,4}$ because of the higher weight of the left sub-divide relation.
- *Optional sub-divide relation*. This is also an n-ary relation allowing optional sub-divisions of a cell providing alternative sub-divisions. For example, a segmentation can be performed by using region growing or optionally by split and merge. For the example in Fig. 4, the cell $C_{1,1}$ can be sub-divided into either $C_{2,2}$ or $C_{2,3}$. The relation is again provided with weights. If there are optional sub-divisions that have to be performed together, this is denoted by an arc in the graph. In Fig. 4, an optional sub-division of cell $C_{2,3}$ may result in $C_{3,3}$ or optionally in $C_{3,4}$ and $C_{3,5}$.
- *Detail relation*. This is a relation between a cell and a detail of the cell. For any cell there may be a number of different possible details. For example, a cell representing edge detection could be detailed in Roberts, Sobel or Canny algorithm.

An example for an analysis graph is given in Fig. 5 on the left, which shows that, for example, rectangle detection can be optionally sub-divided into the cell “Grouping B”, and optionally also into “Grouping A”. Next, “Grouping B” can be optionally sub-divided into the cell “Line detection” or optionally into the cell “Thresholding” which is sub-divided

into “Edge detection”. Up to now, only sub-division relations were encountered in this graph, all of them represent a certain class of image-processing algorithm. Detail relations like the three possible cells “Roberts edge detection”, “Sobel edge detection”, and “Canny edge detection” of “Edge detection” in the graph represent actual operations on data. One possible solution, a so-called instantiation of the analysis graph, would be the rectangle detection using “Grouping B”, “Line detection”, and “Burns line detection”, shown in Fig. 5 on the right.

Each cell of the analysis graph can have a set of in- and outputs, which can be connected to the in- and outputs of other cells in the analysis graph, representing the data flow in the analysis graph (see Fig. 6). The hierarchy in the graph is kept, the output of one cell in one level may be interconnected to an input on the same level, the in- and outputs of cells on different levels represent the same values. In Fig. 6, the input parameters p_1, p_2, p_3 of cell $C_{1,1}$ are either the input parameters for $C_{2,1}$ (p_2) and $C_{2,2}$ (p_1, p_3) or optional input for $C_{2,3}$ (p_1, p_2, p_3). The parameters p_1 and p_3 are computed either by cell $C_{3,1}$ or $C_{3,2}$ represented by the detail relation of cell $C_{2,2}$. Detail relations of $C_{2,1}$ and $C_{2,3}$ are not shown in Fig. 6.

The non-instantiated analysis graph has the benefit that cells can be added and removed without any effect to the rest of the graph. Therefore, as many detection algorithms as available can be tested in parallel, the evaluation of the detection result is performed by the instantiation, all cells that are not necessary for the final solution are eliminated. If the result of the recognition is not satisfying, new cells are added in the original analysis graph which is then re-instantiated to provide a better solution.

The analysis process contributes two elements to the final inspection process, i.e., the building of the description with the corresponding weights for each node, and the analysis graph containing the specific detection order and the specific detection algorithm parameters.

The building up of the analysis graph for inspection can be generalized, since inspection differs only in the description and detection, the overall process is common to all of the inspection problems and consists of four major steps. Figure 7 shows the general analysis graph for inspection. The input of the *object inspection* graph is the *description*, the output is the *result* of inspection. The object inspection can be (non-optional) sub-divided into:

- *image acquisition*. This cell has the image and its statistical parameters; like noise distribution; as the output.
- *Feature determination*. The feature determination determines specific parameters of features (like position and size) within the image. Depending on the image acquisition, feature determination can be sub-divided into:
 - *object detection*. The object has to be located within the image and its size has to be determined, and
 - *feature detection*. Within the object, features used for inspection have to be located.

Optionally, if the image acquisition parameters are fixed (the object has always approximately the same position, orientation, and size), these two cells can be replaced by:

- *hypotheses generation*. Hypotheses about the position, orientation, and size of all features are generated with the help of the description.

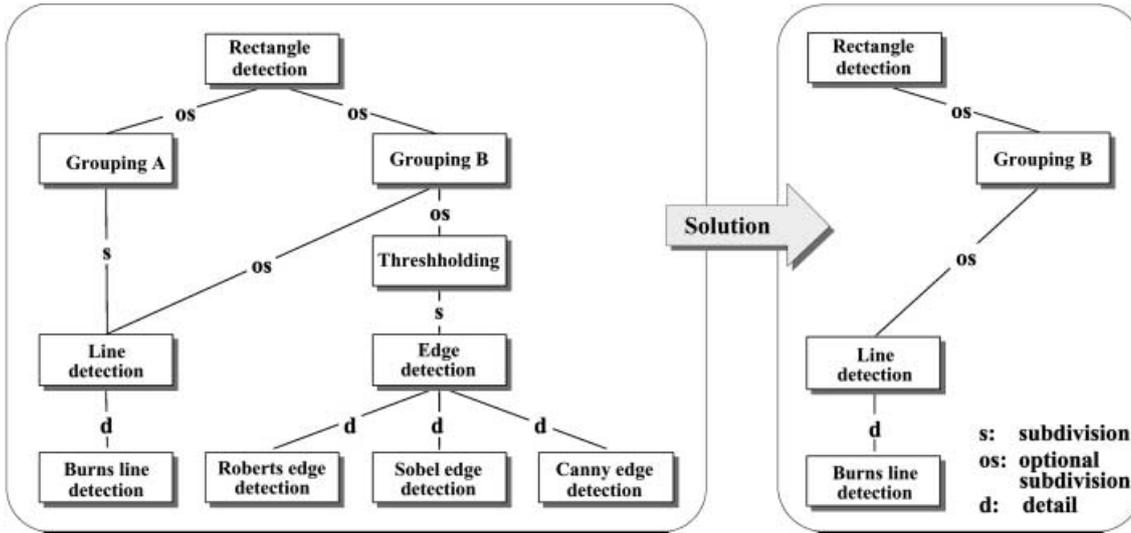


Fig. 5. Example for a hierarchical analysis graph

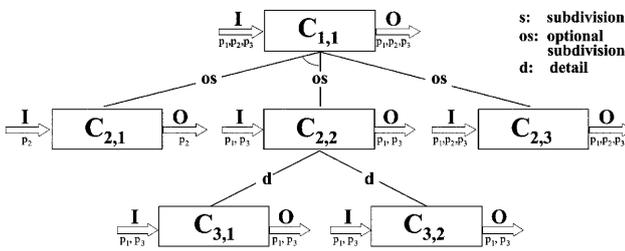


Fig. 6. Data flow in the analysis graph

- *Hypotheses verification.* The specific parameters of the features, either determined by detection or by hypotheses generation, are checked whether they match with the description, using features which were not used up to this step.
- *Model matching.* The final step of analysis matches the actual parameters of the features with the description, thus producing the result of inspection.

The primitives have to be detected in all of the images of the test series. The weights of the analysis graph are set in accordance with their probability of detection. The inspection process “learns” which detection algorithms should be used in a specific area of the image and how many parameters must be found in the image rather than in the description. This introduces a detection verification concept guided by weights in the analysis graph. If some primitives can be found accurately with a high probability, they get a high weight. Primitives with high weights are searched first, if they can be found, primitives with lower detection probability are predicted in a specific part of the image (a hypothesis generation), and then checked for their presence (a verification of the hypothesis).

Evaluation of the cells has to prove that the component is a legal solution to the given problem. In other words, evaluation becomes verification. Experience reveals that applying algorithms with standard parameter values often results in very poor performance in this respect: for example, filtering algorithms need to know the statistical properties of the noise to function properly [11]. One solution to this problem con-

sists in adjusting the parameters interactively, i.e., the user subjectively judges the results on an output device (like the monitor) [22]. The interactive solution defeats the purpose of automated inspection, but it is the only one used so far. Since the human eye is a poor judge for feature detection results, researchers have developed formal error measures to assess the degree of conformance of operator evaluation and segmentation. Furthermore, systems are under development that solve the problem of algorithm selection and parameter adjustment (see, for instance, [3,6–8,33]), while others try to evaluate vision modules and their performance (see [13,16]). However, up to now, no general-purpose image-processing system has been presented because this problem is not trivial to solve [8,9]. We performed selection, parameter adjustment, and evaluation of detection algorithms interactively due to the lack of automated systems in this area.

The instantiation of the analysis graph can also be seen as: given a directed graph, find an optimal solution path with minimum amount of computational effort. For solving this kind of problem, standard techniques (like the A* Algorithm [17,26]) can be used to find the minimum cost path in the graph, i.e., the best analysis strategy.

5 Analysis graph for analogue display instruments

This section shows the applicability of the general analysis graph in the case study of analogue display instruments (ADI). This type of instrument serves as a demonstration, since there are various different types of measuring instruments with innumerable different displays and layouts, but all of them have certain common properties which can be used to build up a specific description. Three common primitives describe analogue measuring instruments (see Fig. 8).

- *Pointer.* A pointer can have any symmetric shape such as line, triangle, rectangle, or a combination of them. In addition, pointers that rotate have a circle at their center of rotation (see Fig. 8). The shape is defined by a primitive, a combination of primitives, or in the case of a shape that is not easily represented by primitives, by a

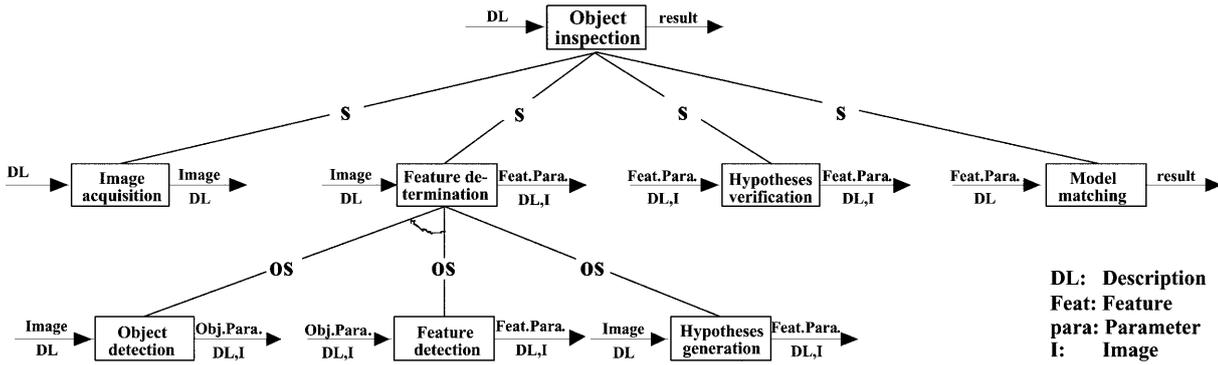


Fig. 7. General analysis graph for AVI

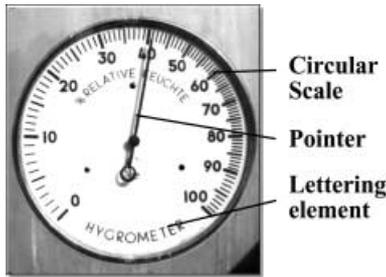


Fig. 8. Primitives of a hygrometer

bitmap, containing one half of the shape and the medial axis.

- *Scale.* The shape of a scale depends on the motion of the pointer; scales with rotating pointers have the shape of a *circle* or a *circular arc*. Pointers moving straight have rectangular scales. Scale captions are considered to be part of the scale.
- *Lettering element.* Such an element carries information about the measurement and the orientation, this includes all writings such as unit, company name, and emblem of maker.

Following the definition of the description, the analysis graph is constructed by applying the general analysis graph to the specific problem. In- and output parameters are described at the top level of the graph only, to simplify the graph. This general concept is used for the definition of the analysis graph for ADIs, resulting in the graph shown in Fig. 9. There are six mayor analysis steps which are the leaves of the graph. If steps object and feature detection are optionally replaced by hypotheses generation the number of steps is reduced to five.

To get an impression of the data flow, Fig. 10 shows the data flow between the leaves of the graph in Fig. 9. Further sub-divisions of individual nodes are shown. Since the graph will not be instantiated for a specific instrument, detail relations are not shown. In the following, the nodes of the analysis graph for ADIs are described.

- *Image acquisition.* To simplify the analysis, we assume that the image contains only a single object [31]. Prior to further computation, the actual image has to be checked whether it satisfies the assumption. This includes a control of the contrast and the histogram. If there are strong variations to the statistical mean image of the inspection

series, there is an error in the illumination, acquisition, or positioning, and the analysis is stopped. The *image* of the ADI in x, y coordinates (rows and lines in image coordinates) is the output of this node. This image is the input for the measuring instrument detection, or optionally for the hypothesis generation if imaging parameters are fixed. Note that only newly determined or computed parameters are explicitly mentioned, all others are implicitly available, like the parameters of the description which are accessible to all nodes of the graph since they are the input of the root of the graph.

- *Measuring instrument detection.* If position and size of the measuring instrument in the image depend on actual image acquisition parameters and positioning of the instrument, this step has to be carried out. The shape of the instrument is looked for in the image with regard to topological, radiometrical, and geometrical features. Depending on the shape, the node is optionally sub-divided into the detection of the geometrical shape like rectangle detection, circle detection, and freeform detection. The leaves of the analysis graph (Fig. 10) are the defined interface to feature detection algorithms. All of the nodes are sub-divided into different possible detection algorithms and detail relations. In the sub-graph *rectangle detection*, for instance, all algorithms to detect rectangles available to the system are included (as in Fig. 4, for example), further ones can be added. The *origin*(x, y) in image coordinates of the object-centered coordinate system and the *size* (in image and world coordinates) of the detected object are the outputs of this node.
- *Measuring unit detection.* Detection and localization of measuring units are carried out in the limited area within the measuring instrument, defined by the hypothesis constructed by the measuring instrument detection. First, the $1..n$ scales are looked for, because they cover a larger area in the image than pointers and their position does not change within the instrument. Note that the search space for the corresponding pointer can be restricted by the region defined by the detected scale. Furthermore, the measuring unit detection may be sub-divided into $0..m$ lettering element detections. This is necessary to determine the orientation of the scale if, for instance, there is only one scale on the instrument. This node supplies the specific origin and orientation for up to n scales and up to m lettering elements.

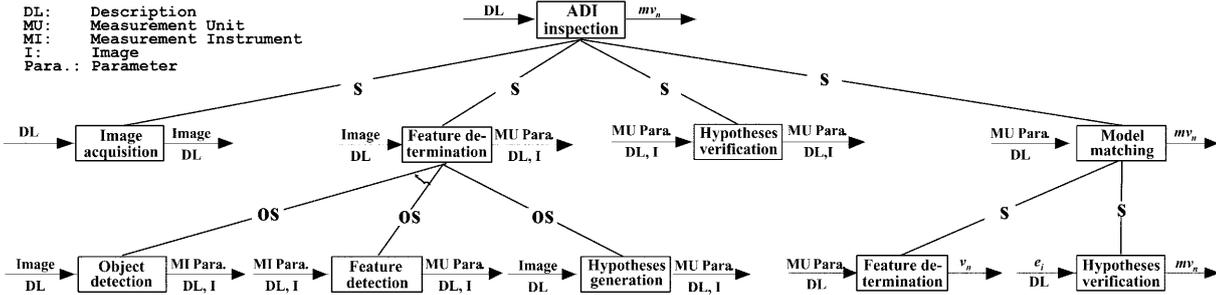


Fig. 9. General analysis graph for ADIs

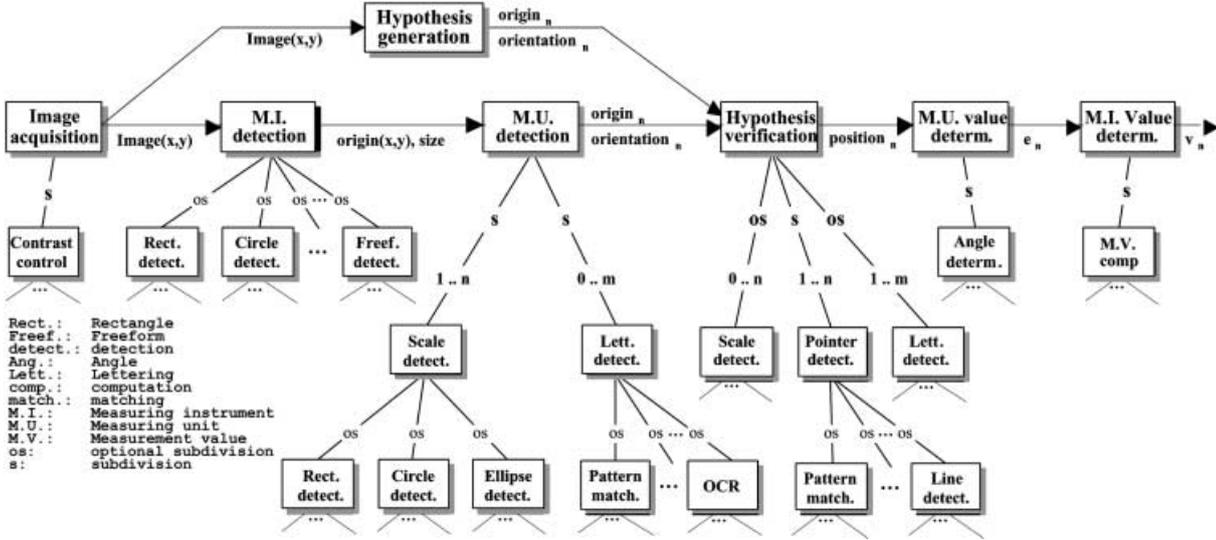


Fig. 10. Analysis graph for ADIs (no detail relations shown)

- *Hypotheses generation.* If the image acquisition parameters are fixed, hypotheses about specific parameters of the measuring instrument and measuring units are generated without time-consuming search in the image. Scales and lettering elements are induced to be on a specific position with a specific orientation within the image. The output is the same as that of the unit detection.
- *Hypothesis verification.* For both sub-divisions of feature determination, a verification of the generated hypotheses is necessary. In order to answer the question “Are the measuring units on the right places on the measuring instrument?”, the identified type is verified by checking the induced position of the lettering elements (and possibly scales) in the image. The aim of the verification is to find out which of the candidates are elements of a measuring unit and which are not. Furthermore, all primitives which were not detected in the previous nodes are detected. In the specific case of ADIs this detection includes pointer detection. The origin of the pointer is the center of the scale, the search space is limited by the corresponding scale. The possibly corrected positions and orientations of the scales and the specific positions of the pointers are the output of this node.
- *Measuring unit value determination.* The value for each measuring unit is determined. The value $e_i, i = 1..n$, for each measuring unit is the result of this node of the analysis graph.

- *Measurement value determination.* The last element of data flow in the analysis graph is the determination of the value mv_n .

The analysis graph (number of nodes after instantiation: $2 \times (4 + 2 \times (n + m))$) describes the analysis for any ADI. In the example of the hygrometer, an instantiation of the analysis graph (24 nodes) is shown in Fig. 11, the imaging parameters are supposed to be unknown. Therefore, the analysis graph of the hygrometer has two sub-divisions of *feature determination*. First, the hygrometer is detected using a circle detection, defining the origin of the object-centered coordinate system. Since the origin of the scale is the same as the origin of the instrument, only one lettering element has to be looked for to determine the orientation of the scales (in this example, the lettering element “left dot” is looked for, using again a circle detection technique along a circle in a known distance). To verify the result, the lettering element “hygrometer” is checked by *hypotheses verification*, i.e., using an OCR algorithm.

If the test is successful, the positions of the pointer is determined by a line detection algorithm. Since the orientation of the scale and the position of the pointer are known, the measurement unit determination, computing the angle between the origin of the scale and the pointer, is performed. The measuring instrument value determination has the result $mv_n = 41$. Figure 12 shows the result for the test image; a humidity of 41% was the correct result.

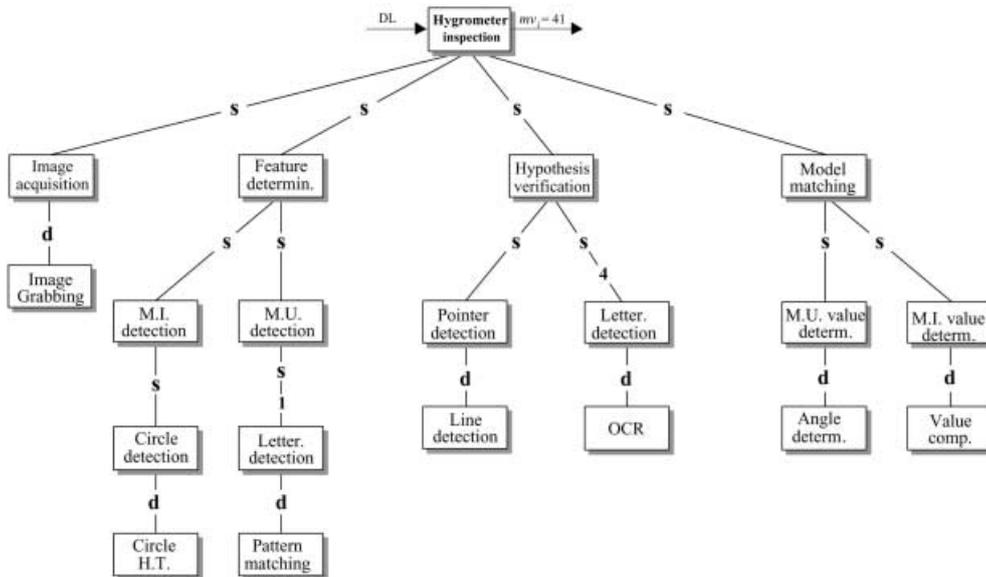


Fig. 11. Analysis graph for hygrometer inspection

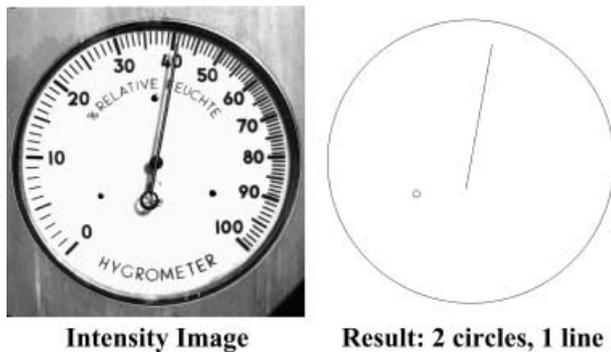


Fig. 12. Result for a hygrometer ($mv_n = 41\%$ humidity)

6 Analysis graph for watermeter calibration

In this section, an industrial application demonstrates the use of the analysis graph, resulting in a successful working inspection system for watermeters, a specific analogue instrument (for details on the calibration process, see [32]). In the specific case of watermeters, there are a type-dependent number of circular scales (two to five) with coupled pointers, one rectangular scale (not on all types), and different types of lettering elements (see Fig. 13). Of these, two are always present and of importance: the serial number, which identifies the watermeter unambiguously, and the rotary counter, which displays the measurement value for further reading in the households. For every type, a description (approx. 20 nodes) is constructed.

In the general case, the orientation and position of the instrument is not known exactly, since positioning errors in the image acquisition may occur, even if the distance to the object is fixed. Therefore, the general analysis graph for watermeters allows a handling of the misorientation and does not only generate hypotheses that are verified afterwards. The analysis graph is type independent; however, there is an interaction with the description of the type. Therefore, it can distinguish, for instance, whether three or four circular scales have to be detected.

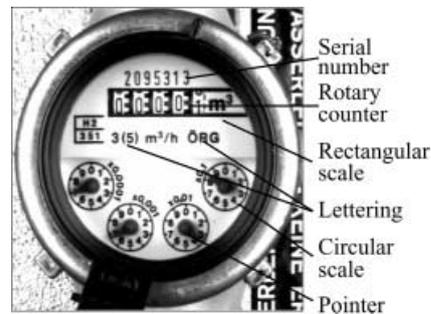


Fig. 13. Primitives of a watermeter

Figure 14 shows the instantiated analysis graph for watermeters with four circular scales. It has 46 nodes, of which 20 nodes are leaves represented by detail relations. To give an impression of the size of the non-instantiated analysis graph, we estimate the number of nodes, since this is dependent on the number of different detection algorithms available at set-up time. Supposing two or three different detection algorithms available for each primitive multiplies the number of optional and detail relations in the graph to approx. 100 nodes for this example.

Following the image acquisition, the position of the watermeter has to be detected within the image. All of the scales (rectangular and circular) have to be detected within the image, since the orientation of the watermeter is not known. The result of the unit detection is checked with the description, performing a hypotheses verification. In case many different types have to be inspected, all three lettering elements (approval sign, nominal performance, and country type sign) which allow a distinction have to be checked. Furthermore, all positions of the pointers have to be detected within the scales, once the hypotheses have been verified. To achieve the complete measurement value, the rotary counter has to be analyzed, and the serial number has to be read to identify uniquely the watermeter under inspection. At this stage, all the generic parameters of the watermeter primitives are known. The model matching consists of a determination of

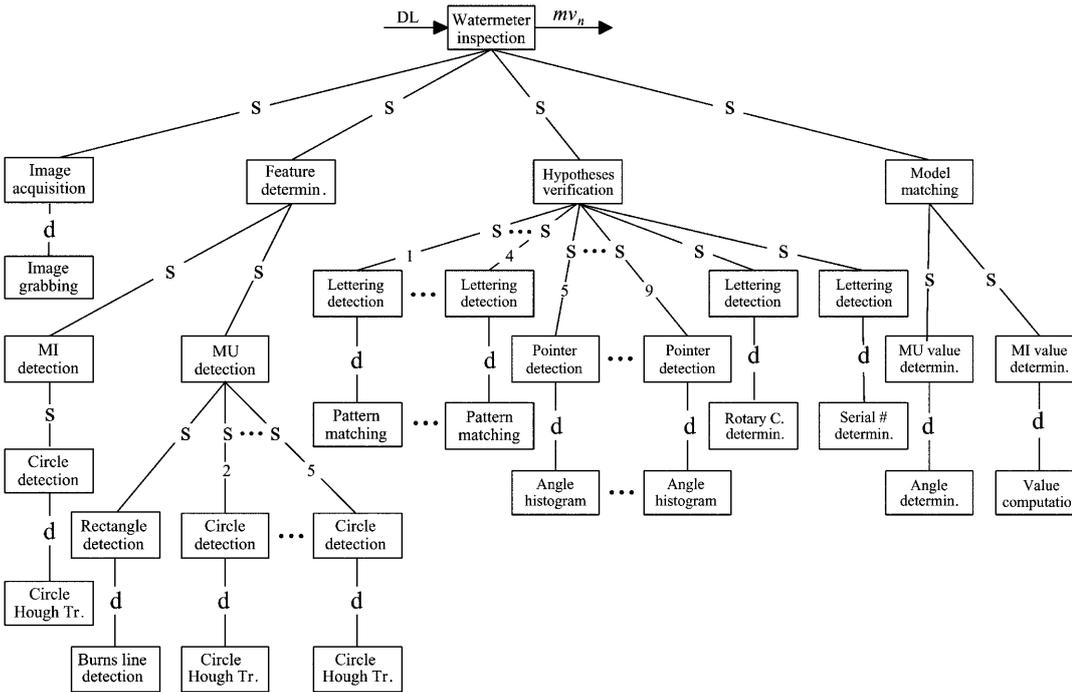


Fig. 14. Analysis graph for watermeter

the value that the single units displays and a value computation regarding the coupled pointers.

7 Results

An inspection system test checks whether the inspection works correctly. The watermeter inspection process was tested comprehensively with a PC configuration [30]. Results of inspection system test:

- number of images in test series: 200 images
- image size: 300x300 pixels
- detection rate: 100%
- accuracy: 100%
- reliability: 95%
- computation time: app. 2 s.

The reliability of a little more than 95% is due to slight changes in the illumination during acquisition and air bubbles inside the watermeter, which influence the pointer and scale detection. Although there was a rejection rate of 5%, the reading accuracy was 100%, because error detection worked properly, and non-readable images were marked and stored for inspection by the operator. All of the computed measurements were correct. Problems with air bubbles disturbing detection occur with rotary counter and serial number determination.

Air bubbles disturb the pointer detection significantly. Therefore, strategies for better detection had to be considered. Instead of using an algorithm that defines the position of the pointer with the help of an angle histogram, a matched-filter approach was used. The exchange of the detection algorithm caused a change of detail relations in the analysis graph, no other adaption had to be made. The test

series with the 200 images was performed again, now attaining a reliability of more than 99% for the measurement value. Figure 15 shows the user interface for a complete calibration run, consisting of an initial value (A:1) and five consecutive measurements for one watermeter. Below the images the computed measurement values are shown (bottom of Fig. 15).

The application of watermeter calibration has to handle an unlimited number of different types. The flexibility within the specific analogue instrument is provided by the description language that handles different layouts by different descriptions. The analysis graph is independent of any change of the description. Another example shows that the description language and the analysis graph, together with the appropriate detection algorithms, can also handle another type of analogue display instrument. With the description of the clock, generated by generic detection and interactive definition of the primitives (three circular overlaid scales, four lettering elements, and three different pointers), the analysis produced the result shown in Fig. 16. In our test series (20 samples, without the specific case of pointers being completely overlaid), all pointer positions were computed exactly, resulting in correct reading of the time in all images.

The working time for adapting the general description and the general analysis graph interactively was approx. 6 h for the watermeter and 1.5 h for the clock. The reason for the relatively long adaptation time for watermeters lies in the explicit formulation of coupled pointers. To solve the problem of overlapping pointers, the measurement unit value determination was changed so that the minutes hand was searched first. If the other hands could not be detected, they were assumed to be overlaid by the minutes hand.

These two examples showed that an inspection system can be set up in a short time for any analogue instrument.

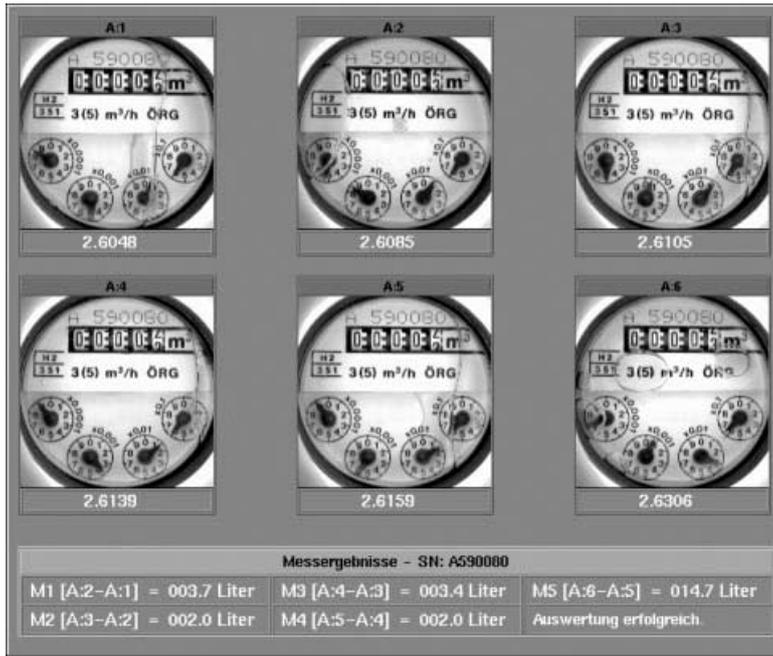


Fig. 15. User interface with computed measurement values

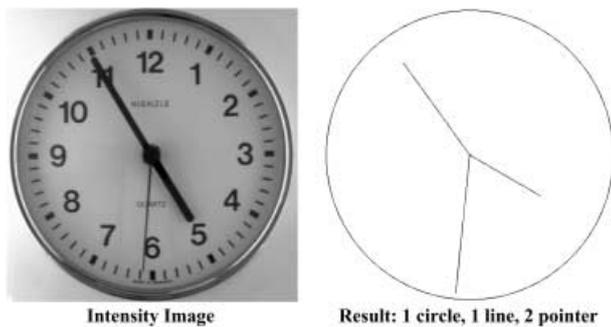


Fig. 16. Result for a clock ($mv_n = 4:54:31h$)

To demonstrate that the general analysis graph does not only work for analogue instruments, a floppy drive casing inspection was simulated. For the casing, it was supposed that six drill holes and one punch had to be inspected using the concept. Using the binary silhouette, the generic detection produced nine circles. Two of the circles were removed from the description and two rectangles were added interactively.

The general analysis graph was adapted for the casing inspection, supposing a nearly fixed orientation of the casing within the image. The model matching was supposed to be a checking for existence. The result indicated whether all of the drill holes were present or not. Figure 17 shows the result after having performed the inspection for a correct casing (Fig. 17a) and a (simulated) defect casing (Fig. 17b). Since one circle in the defect image was not detected, this image was classified as “fault”. The same analysis graph (including detection algorithms) can be used for the inspection of the object using an intensity image, since the parameters of the features do not change. Furthermore, misregistrations do not affect the inspection, since relations between primitives allow their detection.

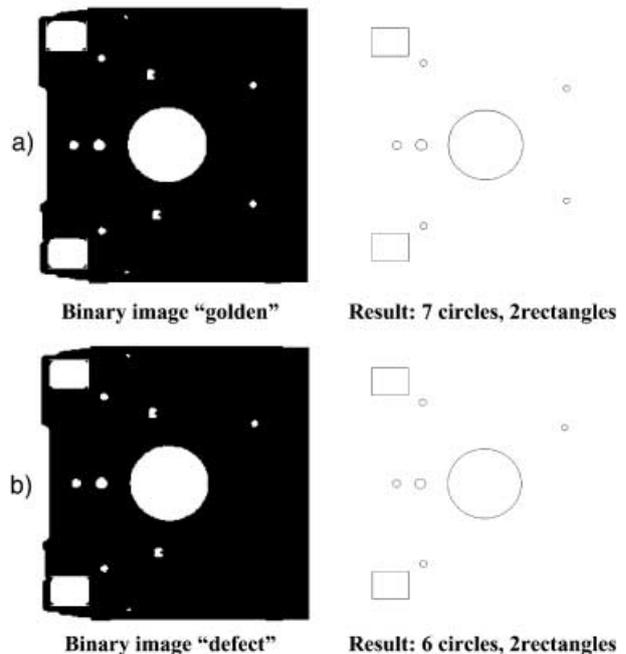


Fig. 17a,b. Casing: a all holes present b one hole missing

8 Conclusion

In this paper, a general **analysis graph** for inspection was presented, where detail relations were used to represent detection algorithms. In the preliminary analysis graph, it was not specified which of the algorithms should be used for the specific inspection, the selection of the algorithm was postponed to a test series. The use of any detection algorithm was possible by changing the analysis graph instantiation in the detail relation, the overall analysis process remained the same. Together with an object-specific description, the analysis graph was instantiated. This systematic approach to

inspection allows its application to a wide range of inspection problems. It can be seen as a “recipe” for solving industrial applications, stating at which stage which kind of decisions have to be made. The systematic approach also permits a high degree of flexibility, since it contains application-specific and application independent parts.

The applicability of the analysis graph was demonstrated on the case study of ADIs. This type of object served as a demonstration, since there were various different types of measuring instruments with innumerable different displays and layouts, but all of them had certain common properties, which were used to build up a specific description. With the help of the description and the proposed general analysis graph, the specific analysis graph for ADIs was defined. Examples demonstrated that this ADI analysis graph could be used to model-specific instruments like hygrometers.

Subsequently, the analysis graph was applied to the calibration of analogue watermeters, where industrial constraints had to be fulfilled. The inspection determined the indicated measurement value and the serial number, protocolling each measurement step. In a test series performed with 200 frames, the positions of **all primitives** were determined in the **requested time** (approx. 2 s), with the **requested accuracy** (100%) and **reliability** (95%). The reliability of a little more than 95% was due to air bubbles inside the watermeter. To overcome this problem, the detection algorithm was changed without changing the analysis. The final test series with 800 images was performed, attaining a reliability of more than 99% for the measurement value.

The flexibility of the analysis graph was demonstrated by testing the analysis process for hygrometers with the description of other ADIs (watermeters and a clock), which was performed by adapting the analysis graph, but without changing the detection algorithms. Since the detection is represented as detail relation in the analysis graph, having a designed interface for primitives, a change of the detection algorithm is possible without changing the overall analysis.

References

- Amerinex Artificial Intelligence Inc. (1995) The KBVision system – User’s Guide. Amerinex Artificial Intelligence Inc., 409 Main Street, Amherst, MA 01002, USA
- Batchelor BG, Whelan PF (1996) Machine vision systems: Proverbs, principles, prejudices and priorities. In: Proc. of SME Conf. on Applied Machine Vision, pp 7–19,
- Bodington R (1995) A software environment for the automatic configuration of inspection systems. In: Proc. of 1st Int. Workshop on Knowledge-based Systems for the (Re)Use of Program Libraries, volume 1, Sophia Antipolis, France, pp 100–108
- Chin RT (1988) Automated visual inspection: 1981 to 1987. *Comput Vision Graphics Image Process: Image Understanding* 41(3): 346–381
- Chin RT, Harlow CA (1982) Automated visual inspection: A survey. *IEEE Trans Pattern Anal Mach Intell* 4(6): 557–573
- Clement V, Thonnat M (1989) Handling knowledge in image processing libraries to build automatic systems. In: Proc. of Int. Workshop on Machine Intelligence and Vision, pp 187–192
- Clement V, Thonnat M (1993) Ocapi: A knowledge-based approach to integration of image processing procedures. *Comput Vision Graphics Image Process: Image Understanding* 57(2): 166–184
- Clouard R, Elmoataz A, Porquet C, Revenu M (1999) Borg: A knowledge-based system for automatic generation of image processing programs. *IEEE Trans Pattern Anal Mach Intell* 21(2): 128–144
- Clouard R, Porquet C, Elmoataz A, Revenu M (1995) Why building knowledge-based image segmentation systems is so difficult. In: Proc. of 1st Int. Workshop on Knowledge-based Systems for the (Re)Use of Program Libraries, volume 1, Sophia Antipolis, France, pp 138–148
- Cowan CK, Bergman A (1989) Determining the camera and light source location for a visual task. In: Proc. of IEEE Int. Conf. on Robotics and Automation, volume 1, pp 509–514
- Crevier D (1993) Expert systems as design aids for artificial vision systems: A survey. *Proc SPIE* 2055: 84–96
- Darwish AM, Jain AK (1988) A rule-based approach for visual pattern inspection. *IEEE Trans Pattern Anal Mach Intell* 10(1): 56–58
- De Boer C, Smeulders A (1996) BESSI: An experimentation system for vision module evaluation. In: Proc. of 13th. Int. Conference on Pattern Recognition, Vienna, Austria, volume 3, pp 109–113
- Even S (1979) *Graph Algorithms*. Computer Science Press, Rockville, Md.
- Fu KS (1982) *Syntactic Pattern Recognition and Applications*. Prentice-Hall, Englewood Cliffs, N.J.
- Haralick RM (1994) Performance characterization in computer vision. *Comput Vision Graphics Image Process: Image Understanding* 60(2): 245–249
- Hart PE, Nilsson NJ, Raphael B (1968) A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans Syst Sci Cybern* 4(2): 100–107
- International Organization for Standardization. (1994) ISO 9000: International Standards for Quality Management, International Organization for Standardization, Geneva
- Khoral Research Inc., USA. (1995) KHOROS Program Services – User’s Guide
- Matrox Electronic Systems Ltd. (1995) *Matrox Inspector – Product Overview*. Matrox Electronic Systems Ltd., 1025 St. Regis Blvd., Dorval, Quebec, H4P 2T4, Canada
- Matrox Image Series – Product Overview (1992) Matrox (UK) Ltd., 6 Cherry Orchard West, Kembrey Park, Swindon, SN2 6UP, UK
- Matsuyama T (1989) Expert systems for image processing: Knowledge-based composition of image analysis processes. *Comput Vision Graphics Image Process: Image Understanding* 48(1): 22–49
- Misra S, Bennett CA (1983) Lighting for visual inspection task. *Light Des Appl* 13(1): 32–33
- Nalwa VS (1993) *A Guided Tour of Computer Vision*. Addison-Wesley, Bonn
- Newman TS, Jain AK (1995) A survey of automated visual inspection. *Comput Vision Image Understanding* 61(2): 231–262
- Pearl J (1984) *Heuristics*. Addison-Wesley, Reading, Mass.
- Peters ED (1994) *ISO 9000 Almanac 1994–1995*. Irwin, Homewood
- Sablatnig R, Menard C (1998) On representing the object model for automatic visual inspection using a description language. *Proc SPIE* 3306: 131–140
- Sablatnig R (1996) Flexible automatic visual inspection based on the separation of detection and analysis. In: Proc. of 13th. Int. Conference on Pattern Recognition, Vienna, Austria, volume III, pp 944–948. IEEE Computer Society Press, New York
- Sablatnig R (1997) *A Highly Adaptable Concept for Visual Inspection*. PhD thesis, TU-Vienna, Inst. f. Automation, Pattern Recognition and Image Processing Group, Vienna, Austria
- Sablatnig R, Kropatsch WG (1994) Application constraints in the design of an automatic reading device for analog display instruments. In: 2nd IEEE Workshop on Applications of Computer Vision – Sarasota Fla, pp 205–212. IEEE Computer Society Press, New York
- Sablatnig R, Leonardi A (1995) Visual inspection of watermeters used for automatic calibration. In: Chin RT, Ip HHS, Naiman AC, and Pong T-C (eds) *Image Analysis Applications and Computer Graphics*, volume LNCS 1024, Springer Verlag, Heidelberg, pp 518–520
- Thonnat M, Moisan S (1995) Knowledge-based systems for program supervision. In: Proc. of 1st Int. Workshop on Knowledge-based Systems for the (Re)Use of Program Libraries, volume 1, Sophia Antipolis, France, pp 4–8
- Young M, Argiro D, Kubica S (1995) Cantata: Visual programming environment for the khoros system. *Comput Graphics* 29(2): 22–24

Chapter 3

Application Constraints in the Design of an Automatic Reading Device for Analog Display Instruments (“Inspection”)

Robert Sablatnig and Walter G. Kropatsch. Application Constraints in the Design of an Automatic Reading Device for Analog Display Instruments. In *Proc. of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota*, pages 205–212. IEEE-Computer Society Press, 1994.

Application Constraints in the Design of an Automatic Reading Device for Analog Display Instruments*

Robert Sablatnig and Walter G. Kropatsch

Institute for Automation, Department of Pattern Recognition and Image Processing,
Technical University Vienna, Treitlstr3/1832, A-1040 Vienna, e-mail: sab@prip.tuwien.ac.at

Abstract

An analysis system design based on experience with a successful application in the field of inspection and calibration of an analog display measuring instrument is presented in this paper. First the measuring instrument is divided into its primitiva, defining the a priori known parameter of the primitiva: shape, relative position and size. According to the shape of the primitiva pattern recognition algorithms are used to detect the primitiva in intensity images. These independent detection algorithms are then grouped into a detecting order in respect to efficiency. Following a discussion of the general design of the detecting algorithm, specific constraints of the application and the industrial environment are considered in order to refine the general design to an applicable and efficient device by modifying both hardware and software configuration depending on the given constraints. Finally, results of the implementation of the algorithm and the constructed image acquisition device are discussed.

1 Introduction

Machines with analog control instruments are still used in industrial production lines and these measuring instruments have to be monitored by factory workers instead of by an automated process monitoring system. Furthermore, analog instruments are often used in places where a continuous power supply is not ensured or too expensive. The calibration of analog display instruments can be automated by using an automatic reading mechanism which is both safer and cheaper.

Reading a measuring instrument means detecting the position of scales and pointers in the intensity image to determine the value the measuring instrument displays. The design of an automated reading process for an indus-

trial application requires that we consider not only pattern recognition algorithms for solving the problem but also several constraints which are given by the industrial partner and the industrial working process.

In this paper the general recognition theory for many different types of measuring instruments shown in [8] is extended and described in more detail, considering both industrial constraints and the fact that no redesign of the working process should be necessary if, for example, another measuring instrument is used, or if the pointers have different colors, or if the illumination conditions differ, or if the measuring instrument is rotated or if other changes occur. To reach this goal, the primitiva of the measuring instruments and the application constraints have to be determined (Section 2). According to the shape of the primitiva, pattern recognition algorithms to detect the primitiva are tested and integrated into a specific detection algorithm (section 3). However, before the detection process can take place, the important step of image acquisition and illumination must be carried out (section 4). The final development stage adapts the working process to the constraints given by the industrial environment (section 5). The paper concludes with a discussion of the results.

2 Primitiva and constraints

Pointers, scales and lettering elements are the primitiva of an analog measuring instrument (Fig. 1).

A **pointer** can have any symmetric shape with a medial axis such as a triangle, a rectangle or a combination of both. Pointers that rotate have a circle at their center of rotation. The shape is defined by a bitmap, containing one half of the shape and the medial axis.

The shape of a **scale** depends on the motion of the pointer. The shape of a rotating pointer is a **circle** or a **circular arc**. Pointers moving straight usually have rectangular scales. Scale captions are considered as part of the scale.

* This work was supported by the Austrian National Fonds für die gewerbliche Wirtschaft under grant Z13/9934.

There are other layout elements of a measuring unit that carry information about the measurement and the global orientation, classified as **lettering** (e.g. ÖRG in Fig. 1). This includes all forms of writing such as unit information, company name, firm's symbol, etc.

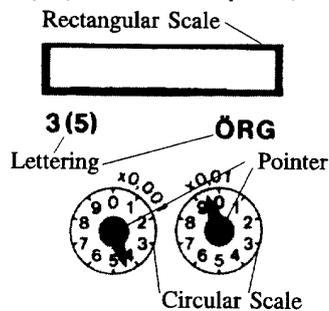


Fig. 1 Elements of a measuring instrument

One constituent is implicitly given with any measuring instrument - the measurement space. It defines what unit and what value the **measuring unit**, consisting of a scale and a pointer, displays and at what sample rate the displayed value can be read. The measuring space also defines the absolute measuring value the **measuring instrument** displays, as a combination of all measuring units. Note that we assume that one measuring instrument can measure only one physical unit.

Besides the primitiva, several constraints are given by the industrial working process and the industrial partner. The most important of these are **computation time**, **accuracy**, **extendability** to other instruments, and **cost**. Therefore it is necessary to find an optimal balance between these constraints.

3 Detection algorithms

In the following we report our experience with several methods for detecting basic primitiva. The kind of elements to be looked for are circles, rectangles and pointers. Therefore, algorithms for detecting these shapes were tested and combined in order to be able to detect the primitiva in the image and to read the value the measuring instrument displays. The primitiva's relative parameters, such as relative size, relative position on the instrument and valence, have to be known a priori to detect them. These parameters form the vocabulary and the relations between the primitiva, the grammar of a description language which can be used by the analysis process to make hypotheses about the position of the primitiva and

to verify them. The general recognition theory for several types of analog measuring instruments is shown in [8].

For an industrial application like quality control in the manufacturing process this general recognition process has to be adapted for a very limited number of different types while taking industrial and economic constraints into account. In this section we describe the algorithms used for primitiva detection and the analysis process used for this specific application where only one type of measuring instrument is used.

3.1 Circular scale detection

To detect circular, arc-shaped scales in the intensity image, we use a detection method based on the Hough transformation [7,10]. The Hough method [6] is extended to circle detection by Duda and Hart [4] and extended by Ballard and Brown by using the gradients [3]. The detection process has the following steps:

Edge detection: The local derivatives in the x and y direction are computed and form the gradient;

Hough transform: Circle centers are clusters in the accumulator space;

Peak enhancement: The accumulator image is convolved with a 9x9 Laplacian-like peak filter as derived in [4];

Max finding: A procedure finds the center of the peak with a neighborhood operation.

The result of circle detection is a set of potential candidates for arcs and circles. The circular Hough transformation estimates the position and the radius of the circular arc or circle. To verify if the candidate is a circular arc or circle, the edge points of the supposed circle are counted along the circle. If the density of detected edge points on a sector of a certain angle is more than 70% of the circle points in the sector, this sector is accepted as a circular arc; if 70% of the circle sectors are accepted, the supposed center is a center of a circle.

Fig. 2 shows the result of detecting circles in intensity images. These circles are the scales of a measuring instrument with 4 circles and all of them were detected.

We carried out a test series with 50 different images of several measuring instruments of the same type. The images contained the measurement instruments (app. 4 inch x 4 inch) in different resolutions (from 100x100 pixels to 400x400 pixels) and under different lighting conditions. Approximately 80% of the circles were detected in the test series. In the images with high resolution (400x400 pixel), 100% of the circles were detected.

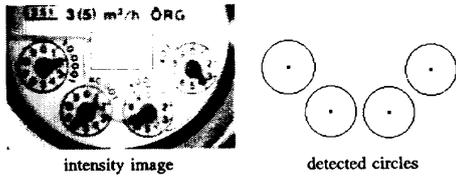


Fig. 2 Detected circles

3.2 Rectangular scale detection

A rectangle consists of 4 straight lines, which we detect using an approach developed by Burns et. al. [2]. This algorithm groups pixels into line support regions of similar gradient orientation. The structure of the associated intensity surface is used to determine the location and properties of the edge. Four steps can be distinguished:

1. Group pixel in line support regions;
2. Approximate the intensity surface by a planar surface;
3. Extract attributes length, location, orientation;
4. Filter lines with certain length, location and orientation;

Steps 3 and 4 were changed from the original algorithm because our aim is to detect lines belonging to rectangles in the intensity image. Only three features given by the Burns algorithm are used, leaving long straight lines in the line set. An orientation histogram is computed and lines with 90° orientation difference are combined to form rectangles (Fig. 3).

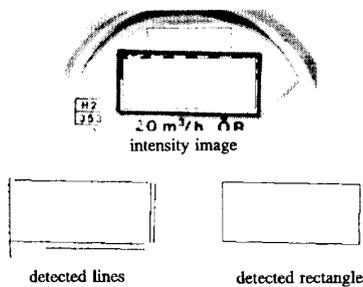


Fig. 3 Detected rectangle

3.3 Pointer detection and verification

We tested two different methods for detecting the pointer in the scale area:

M1. Independent computation of gray level profiles along curves in the image plane: The intensity values are

summarized along an axis and the center is supposed to be the location of the pointer. The mean intensity value of the pointer is computed in the center of the circle. Subsequently this mean intensity value is used to count pixels with similar intensities in the angle histogram.

M2. Multiresolution computation: This method is based on dividing the scale area into sectors and computing the weighted deviation from mean. This deviation gives a probability for the position of the pointer in a sector. Coupled pointers have a special feature: the previous pointer defines the position of the following pointer as shown in Fig.4. The pointer displaying the lower valence (e.g. 1) defines the position of the pointer displaying the higher valence (e.g. 10). If the lower pointer displays 0 for instance, the pointer with the higher valence has only a limited number of possible positions as shown in Fig. 4a. If the lower pointer displays 5 for instance, the pointer with the higher valence has again a different, limited number of possible positions (Fig. 4b). Every measuring unit contributes one digit to the number to be measured. The reading of the values indicated by the pointer position starts with the lowest valence. This strategy is characterized by low error propagation.

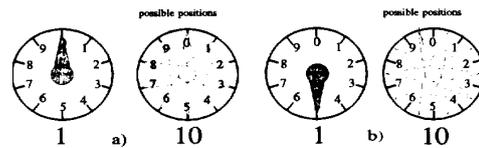


Fig. 4 Reduced search space

Let \bar{e}_i denote the measurement value and e_i the correct value. If the reading error for e_{i+1} is $|\bar{e}_{i+1} - e_{i+1}| < \delta$, the reading error for the next

valence is: $|\bar{e}_i - e_i| = \left| \frac{\bar{e}_{i+1}}{u_{i+1}} - \frac{e_{i+1}}{u_{i+1}} \right| < \frac{\delta}{u_{i+1}}$ For the

unit of the next valence $u_{i+1} > 1$ holds (e.g. 10 in Fig.4), therefore the error of the lower valence has a small propagation and has little influence on measurement determination.

The second method was applied to the special type of measuring instrument used by our industrial partner. Fig. 5 shows the result of the pointer detection. The upper half shows the intensity images of the pointers and the lower half the results of the applied analysis algorithm. Note that the computed value is always the previous value on the scale if the pointer has not yet passed the scale value. For example, one might think the computation has

led to incorrect results, because the rightmost pointer displays the value 2 and the computed value is 1. In reality the pointer has not yet passed the value 2, because the pointer with the lower valence displays 9.

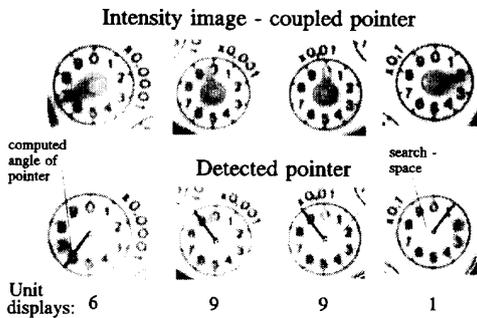


Fig. 5 Detected coupled pointers (0.1996)

3.4 Lettering

The detection of lettering elements has two different purposes. Lettering elements can be taken to check and verify the type and orientation of the measuring instrument. In this case the lettering element is postulated in a certain position and orientation in the image and checked by computing the correlation coefficient between the window found in the intensity image and the bitmap of the lettering element. The computed correlation coefficient defines the similarity between the detected area and the generic area and gives a probability for a match.

3.5 Detection strategy

This section briefly describes the steps of the analysis process. In order to simplify the diagram, Fig. 6 does not show the interaction of all processing steps with the parameters of the a priori given primitives (like size and relative position) and error treatment. If an error occurs in any processing step, the analysis process is stopped.

Image acquisition: Image acquisition is the first important step in the analysis process. Details of hardware and lighting conditions are described in the following section.

Measuring instrument detection: Because position and size of the measuring instrument in the image are not known a priori they have to be detected. The shape of the instrument to be detected (in our case) is circular, so we use circular scale detection for detecting the instrument in the image with regard to topological, radiometrical and

geometrical features of the instrument. The result is the detection of the instrument, its center (origin of the object-centered coordinate system) and its image size.

Scale and pointer detection:

The detection and localization of measuring unit primitiva is carried out in the limited area of the previously detected measuring instrument. First, all circular scales and the rectangular scales are looked for because they cover a larger area in the image than the pointers. Therefore scale detection (both circular and rectangular) is more robust than pointer detection. The search space for the corresponding

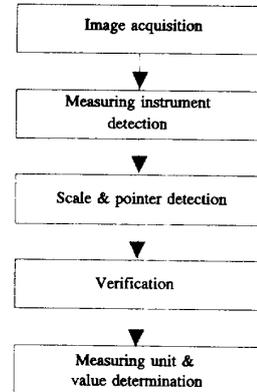


Fig. 6 Analysis process

pointer is restricted by the region defined by the detected scale (section 3.3). This analysis step supplies the specific position and size for all scales and pointers as well as the orientation of the instrument.

Verification: Verification of the generated hypotheses about the size and the orientation of the measuring instrument is necessary. In order to answer the question "Are the measuring units in the right places on the measuring instrument?", the detected measuring instrument is verified by checking the supposed position of the lettering elements in the image. This step also helps to avoid computation of measuring instruments that have similar size and primitiva but were not tested for automatic reading.

Measuring unit and value determination: The value for each measuring unit is determined by relating the position of the pointer to the orientation of the scale. The value for each measuring unit is the result of this processing step. The values of several measuring units are combined according to their valence into the measurement value of the instrument.

4 Image Acquisition and Illumination

One of the most important steps in the analysis process is image acquisition, because the quality of the images strongly influences the quality of the result. In order to ensure an accurate analysis, the image of the object should have high contrast and there should not be any shades or reflections in the image.

4.1 Image acquisition

There were 3 constraints for the image acquisition given by our industrial partner:

- low cost hardware
- pointers can move during image acquisition
- 12 inch maximum distance between camera and measuring instrument

To optimize resolution and the constraint of low hardware costs we used a commercial quality monochrome CCD camera Sony XC 75CE with a resolution of 752 x 582 pixel in connection with an suitable low cost monochrome framegrabber board.

To fulfill the second constraint of "moving pointers", the camera type with shutter option is used. The shutter speed can be selected from 1/125 to 1/10000 seconds. The maximum speed of the pointer is 15 revolutions per second and there are 10 different scale positions. A shutter speed of 1/250 was selected to ensure a minimum accuracy of ± 18 degrees of the position of the pointer in the image at a certain time. At lower pointer speed the accuracy of the pointer position is of course higher.

The maximum distance between measuring instrument and camera was set at 12 inch due to the industrial environment. Therefore we used a 0.3 inch C-mount lens of standard quality for the image acquisition. Lenses with a shorter focal length (0.2 inch and 0.25 inch) were tested but not selected because the distortion of the standard quality lenses was too high. The circular scales on the measuring instrument became elliptical and the circle detection did not operate without lens correction. In order to avoid camera calibration and lens correction an 0.3 inch lens was used for image acquisition.

4.2 Illumination

The industrial environment allows only two solutions for the illumination problem. Either complex preprocessing has to be done to adjust different illumination conditions in the factory or constant illumination is provided by a special illumination device designed for the special illumination problem [1]. We chose the second solution for two reasons: it helps to ensure correct reading and it can be implemented in our special environment.

Like most of the measuring instruments, our instrument is covered with a glass plate. As a result there are specific problems with specular light and shadows. With specular light a reading of the pointers is impossible be-

cause a highlight in the area of a pointer makes the pointer invisible and therefore unreadable. These highlights can be avoided if the angle between the light reflected on the glass plate and the optical axis of the camera is more than 30° as shown in Fig. 7.

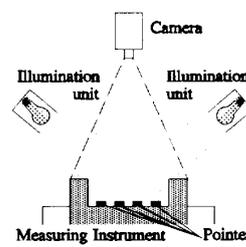


Fig. 7 Illumination avoiding highlights

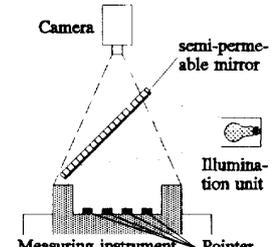


Fig. 8 Profile projection illumination

Another illumination problem that causes incorrect reading of the pointers results from the shadows produced by the pointers. Pointers always have to have some distance to the background in order to be moveable (in our case app. 0.1 inch). This distance produces shadows cast by the pointer and therefore ambiguities in the determination of the pointer's position. A result of this illumination configuration is shown in Fig. 10a where shadows cast by the pointers can be seen. For this reason and because of the special shape of our measuring instrument, the illumination-configuration shown in Fig. 7 cannot be used to illuminate the measuring instrument.

Another commonly used illumination method in industrial applications is called the profile-projection method, which uses a semipermeable mirror positioned at an angle of 45° to the optical axis of the camera to diffract the light on its specular side onto the surface of the object (Fig.9). The illumination direction is at an angle of 90° to the optical axis and is deflected in the direction of the optical axis. Subsequently the light is reflected from the object without producing shadows. The drawback of this configuration is that it works only if there is no glass plate on the measuring instrument. In our case the light was totally reflected by the glass plate as shown in Fig. 10b. Therefore we had to construct another

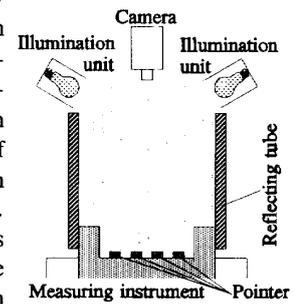


Fig. 9 Reflecting tube illumination configuration

illumination- configuration that overcomes the difficulties with specular lights and shadows. A homogeneous, diffuse illumination of the measuring instrument is necessary to obtain a good basis for further computation.

A tube with a diffuse, specular inner surface was affixed to the measuring instrument as shown in Fig. 9. This 6 inch high reflecting tube was illuminated by two lamps on the top of the tube with an illumination angle of app. 70° with respect to the optical axis of the camera. The surface of the tube refracts the incoming light diffusely onto the opposite side of the tube, there it is refracted once again and so on. The set- up of this illumination configuration turned out to be optimal in our tests because neither highlights nor shadows disturb image acquisition (Fig. 10c).

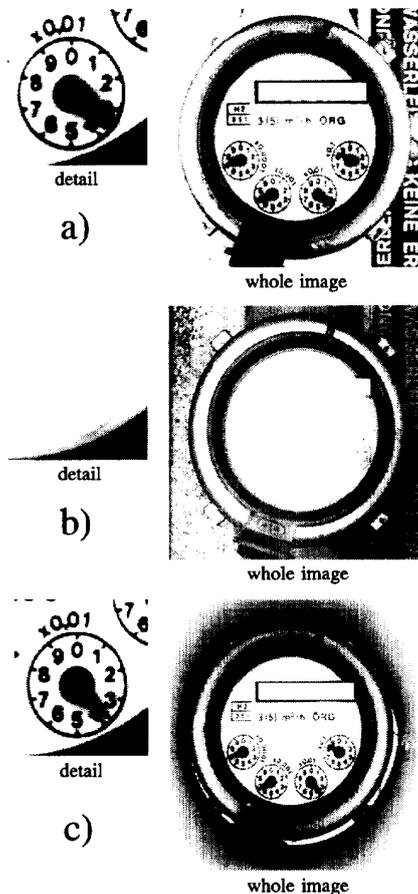


Fig. 10 a) highlight avoiding b) profile projection illumination c) reflecting tube

In our first test series in the industrial environment the tube was used even without the lamps, using the ambient light in the factory. It turned out that a correct result of the automatic reading process is given with a probability of more than 90%.

5 Tradeoffs

The general detection strategy presented in section 3.5 was first implemented in an experimental developing environment on a Unix platform with a series of test images taken under laboratory conditions to take advantage of fast prototyping with high computation power. After demonstrating the practicability of the presented detection strategy, the working process had to be adapted to an industrial environment. Five main constraints have to be fulfilled for this application:

Speed: The reading of the pointer has to be performed in a previously defined time of max. 3 seconds/frame.

Cost: Due to industrial application and reselling strategies, the hardware cost had to be minimal.

Accuracy: Since the automatic reading is used for calibration and industrial process monitoring, reading accuracy plays an important role.

Reliability: The accuracy, the reliability has to be close to 100% to ensure correct calibration and monitoring.

Archivation: In order to fulfill the ISO 9001 quality inspection standard, all measurements have to be archived in image and numerical form together with further data of the measuring process such as time, date, serial number and the like.

The archivation constraints were taken care of by using JPEG data compression for image archiving [9] and a database. The rest of the constraints had to be considered in more detail because the automatic reading process should be integrated into an existing working process, in which the maximum computing time is set at 3 seconds per frame. This computation time was barely reached by using a SUN Sparc 10 workstation for which the hardware costs are already too high for our industrial partner. Therefore we made several tradeoffs between speed, costs, accuracy and reliability in order to achieve an optimal balance of the given constraints.

5.1 Speed versus number of free parameters

The reading process has three image acquisition parameters which allow a very general application of the

reading process to a given measuring instrument:

Size (s): The measuring instrument can have any size within the intensity image (of course there is a minimum size defined by resolution and reading accuracy). The change of the size is caused by the change of the distance between camera and measuring instrument.

Position (x,y): The measuring instrument can have any position in x and y direction within the image.

Orientation (o): The measuring instrument can have any orientation in respect to the camera orientation.

If all of these parameters are not fixed to a certain range, the reading process takes 3.7 seconds. Therefore a successive restriction of these parameters reduces the computation time while decreasing the generality. First the size of the measuring instrument is fixed in the analysis process. In this way, the search space of the circle detection (see section 3.1), detecting the measuring instrument is confined to a small number of possible radii (in respect of δs) and therefore performed with higher speed. This modification saved 20% of computation time and therefore the process fulfilled the time constraint but forced a fixed distance between camera and measuring instrument.

Due to a modification of the final image acquisition configuration it was possible to ensure the position of the measuring instrument on a nearly fixed place within the image, restricting the position parameters to $[x \pm \delta x, y \pm \delta y]$. Therefore a modification of the analysis process could be made by eliminating the processing step "measuring instrument detection", again saving 25% of the computation time. The fixed imaging geometry requires an accurate positioning of the camera or the measuring instrument and was obtained by designing an attachment including illumination units and camera.

5.2 Speed versus costs

Hardware costs play an important role in industrial applications if the system is to be widely used and be sellable. This led us to the decision for a standard PC 486 configuration instead of a workstation configuration, reducing the hardware cost by 80%. This cost reduction has the drawback of dramatically decreasing computation power. After converting the software onto the PC we measured a computation time of the analysis process of 20 seconds/frame, including scale and pointer detection, determination of orientation and measurement value.

As a consequence, the last free image acquisition parameter, orientation, had to be fixed. The fixation of orientation was possible by adapting the attachment device such that it can only be attached in a certain, mechanically fixed way. Subsequently the detection strategy was modified in the following way: only the left- and right-most circular scales are looked for in the intensity image instead of all scales. The size of the search window is defined by the number of pixels representing the positioning error $x \pm \delta x, y \pm \delta y$ and the size of the circular scale (in our case δx and δy are 5% of the diameter of the circular scale). This fixation of the position reduces the search space down to 10% of the previous search area (the complete measuring instrument).

Out of the relative position of the centers of the two detected scales all other scale positions are computed before a simple verification and the pointer detection (see section 3.3) takes place. With this adaption to the detection strategy the computation time is within the desired 3 seconds per frame limit with the requested accuracy.

5.3 Speed versus accuracy and reliability

The tested accuracy of the general detection strategy is $\pm 3^\circ$ determined by the given resolution of app. 500x500 pixels for the measuring instrument. This resulted in a computation time for the adapted reading process (all acquisition parameters fixed) of 4 seconds. Due to the use of coupled pointers with discrete positions and a graduation of 10 units in a circle in combination with low error propagation (see section 3.3), the minimum reading accuracy of the pointer angle detection is $\pm 18^\circ$. This accuracy decreases the reliability and requires very stable illumination conditions in order for the position of the leftmost pointer to be read correctly. Therefore a tradeoff between accuracy and speed has to be made in order to ensure the requested reliability. Our tests demonstrated that an accuracy of $\pm 8^\circ$ is the best tradeoff between speed and accuracy. This reading accuracy reduces the resolution to 300x300 pixels for the measuring instrument and led to a computation time of app. 2 seconds per frame which is within the time constraint.

5.4 Experimental results

The adapted algorithm was tested comprehensively with our PC configuration both under laboratory and industrial conditions. The position of **all elements** except

the lettering was determined in the **requested time**, with the requested **accuracy** and **reliability**.

For the example in Fig. 11, the analysis process computed 0.8361. Note that the coupled pointer computation method is used, therefore pointers which have not passed a value on a scale completely are considered to display the previous value. The analysis process is tested with a test series of 200 frames under laboratory conditions. All of them were computed correctly.

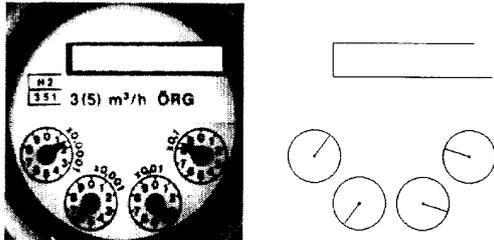


Fig. 11 Detected scales and pointers: 0.8361

During a test series (400 frames) conducted under industrial conditions without special illumination (the test series was conducted with the configuration in Fig. 9 but with ambient light instead of special illumination), the lack of light led to images of very low contrast, causing a rejection rate of 5% (20 frames were not computed because there was not enough contrast in the image to find the circular scales and hence the pointers). Although there was a rejection rate of 5%, the reliability of the reading process was 100% because images that could not be read were marked unreadable and stored for visual inspection by an operator. All of the computed measurements were correct.

6 Conclusion

In this paper we have presented a special application in the field of inspection and calibration of an analog display measuring instrument. A general approach was designed by dividing the measuring instrument into its primitiva and defining the a priori known parameter of the primitiva: shape, relative position, and size. According to the shape of the primitiva suitable pattern recognition algorithms were implemented and tested in order to detect these primitiva in the intensity image. After testing and refining the primitiva detection a general detection strategy was developed to solve the problem of reading the measuring instrument automatically. Instead of detecting primitiva independently, a detecting order was used to

enhance efficiency. The last stage in the design of the analysis process refines the coarse analysis process by adding constraints generated by the specific application and the industrial environment, which are: **speed**, **cost**, **accuracy** and **reliability**.

This refinement influenced both the hardware and software configuration with respect to the given constraints and practicability. Therefore we have suggested a general approach for the solving of a specific problem and have refined this approach by adding constraints given by the specific application. These constraints are heavily interdependent and have therefore to be balanced in an optimal way in order to be applicable in an industrial process.

Acknowledgment

The authors want to thank Hannes Foettinger and Robert Brandner for their help during the adaption of the analysis process for industrial application.

References

- [1] R.J. Ahlers, H.J. Warnecke, "Industrielle Bildverarbeitung", Addison-Wesley Publishing Company, Bonn, Muenchen, New York, Sidney, Tokyo, 1991
- [2] J.B. Burns, A.R. Hanson, E.M. Riseman, "Extracting Straight Lines", *IEEE Transactions on PAMI*, Vol. 8, No. 4, pp. 425-455, 1986.
- [3] Ballard, Brown, "Generalizing the Hough transform to detect arbitrary shapes", *Pattern Recognition* Vol. 13(2), pp. 111-122, 1981.
- [4] P.E. Danielsson, O. Seger, "Rotation Invariance in Gradient and Higher Order Derivate Detectors", *CVGIP*, Vol. 49, pp. 198-221, 1990.
- [5] R.O. Duda, P.E. Hart, "Use of the Hough transform to detect lines and curves in pictures", *Com. of ACM*, Vol. 18(9), pp. 509-517, 1975.
- [6] P.V.C. Hough, "Methods and means for recognizing complex patterns", U.S. Patent 3,069,654, 1962.
- [7] P. Kierkegaard, "A Method for Detection of Circular Arcs Based on the Hough Transform", *Machine Vision and Applications*, Vol. 5, pp. 249-263, 1992.
- [8] R. Sablatnig, W.G. Kropatsch, "Automatic Reading of Analog Display Instruments", accepted at the 12th ICPR, Computer Vision Track, Jerusalem, 1994
- [9] G.K. Wallace, "The JPEG Still Picture Compression Standard", *Com. of ACM*, Vol.34, No.4, pp 30-44, 1991
- [10] H.K. Yuen, J. Princen, J. Illingworth, J. Kittler, "Comparative Study of Hough Transform Methods for Circle Finding", *Image and Vision Computing*, Vol. 8(1), pp. 71-77, 1990.

Chapter 4

Automated Segmentation of Archaeological Profiles for Classification (“Archaeology”)

Martin Kampel and Robert 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.

Automated Segmentation of Archaeological Profiles for Classification *

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

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 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. The one we propose is based on B-splines or bell-shaped splines.

1 Introduction

A large number of ceramic fragments, called sherds, are found at excavations (see Figure 1). These fragments are documented by being photographed, measured, and drawn; then they are classified and stored in boxes and containers. The purpose of *classification* is to get a systematic view on the excavation finds. As the conventional method for documentation is often unsatisfactory [9], we are developing an automated archivation system with respect to archaeological requirements [6], that tries to combine the traditional archaeological classification with new techniques in order to get an objective classification scheme.

A graphic documentation devised 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 2 different illustrators resulting in two different vessels as shown in Figure 2. Note the different shape and decoration, the rim and the thickness for instance are significantly different.

Manual drawings like Figure 2b indicate the beginning

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

and end of defined shape features by horizontal lines. The left half of Figure 2b shows the shape features defined by archaeologists and a subdivision of the profile into intervals. They depict the borders of certain parts of the vessel like rim, wall, and base in this example. By classifying the parts of the profile, the vessel is classified, missing parts may be reconstructed with the expert knowledge of the archaeologist [13]. Segmentation of the profile is done for three reasons: to complete the archive drawing, to classify the vessel and to reconstruct missing profile parts.



Figure 1. Boxes filled with ceramics stored in archives.

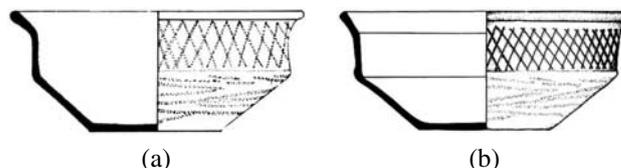


Figure 2. Same vessel drawn by two different illustrators.

Following this manual strategy, the profile should first be segmented into its parts, the so-called *primitives*, automatically. Our approach 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. The segments of the curve are

divided by so called segmentation points. If there is a corner point, that means a point where the curvature changes significantly, the segmentation point is obvious. If there is no corner point, the segmentation point has to be determined mathematically.

The curve is characterized by several points. Figure 3 shows the segmentation scheme of an S-shaped vessel as an example. A set of points is defined like,

inflexion point (IP): point, where the curvature changes its sign;

local maximum (MA): point of vertical tangency;

local minimum (MI): point of vertical tangency;

orifice point (OP): outermost point, where the profile line touches the orifice plane;

base point (BP): outermost point, where the profile line touches the base plane;

point of the axis of rotation (RP): point, where the profile line touches the axis of rotation.

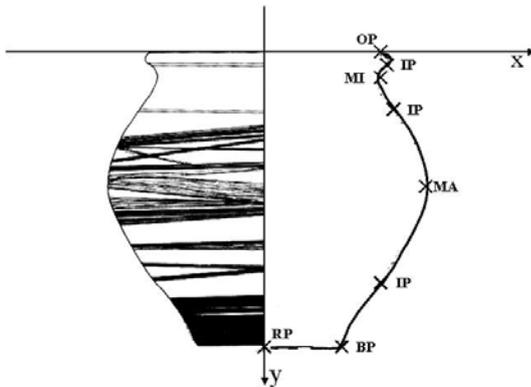


Figure 3. S-shaped vessel: profile segmentation scheme.

By means of these curve points several main segments of a vessel are distinguished: rim, upper part, lower part, neck, shoulder, belly and bottom. On the basis of the number and characteristics of these segments different kinds of vessels can be classified.

2 Automated Segmentation

The profile sections are achieved automatically by a 3D-measurement system based on structured light and a two laser-technique [7]. The profile determined has to be converted into a parameterized curve [12, 5] and the curvature has to be computed [2, 8]. Local changes in curvature [11] are the basis for rules required for segmenting the profile.

Our formalized approach uses mathematical curves to describe the shapes of the vessels and their parts. The profile is thus converted into one or more mathematical curves. We apply four methods for interpolation and four methods for approximation by B -splines on the reconstruction of the

vessel profiles (i.e. the profiles are projected into the plane).

2.1 Interpolation by Cubic Splines

The following definitions were adopted from [4]. We suppose that the planar closed curve r to be fitted (interpolated or approximated) will be represented by parametric equations

$$\mathbf{r}(t) = [\mathbf{x}(t), \mathbf{y}(t)] \quad (1)$$

in an interval in the Cartesian coordinates of \mathcal{R}^2 and has continuous second derivatives. The curve is given by a set of points $P_i = [x(t), y(t)]$ together with the non decreasing sequence of knots $\{t_i, i = 1, \dots, n + 1\}$ of parameter t . Constructing a curve $S(t)$, which approximates the function given by the points can be done by a cubic spline with an adequate parametrization and external conditions. The curve must be initially divided into sub-intervals, where functional approximation and interpolation methods can be applied.

The support of a cubic spline is 5 intervals. Denote by B_i^4 an k -th order spline ($k \leq 3$) whose support is $[t_i, t_{i+4}]$. Then, it is possible to normalize these splines so that for any $x \in [\phi,]$

$$\sum_{i=-3}^{n+3} B_i^4(x) = 1 \quad (2)$$

Any cubic spline $S_n(x)$ with knots t_0, \dots, t_n and coefficients $a_{-3}, a_{-2}, \dots, a_n$ can be written in the form

$$S_n(x) = \sum_{i=-3}^n a_i B_i^4(x) \quad (3)$$

There are $n + 3$ coefficients a_i in representation (3) showing that the vector space of cubic splines has dimension $n + 3$, so that the $n + 1$ functional values will not determine $S_n(x)$ uniquely - two additional constraints must be supplied. Cardinality of the basis has been sacrificed for small support in the basis. Consequently, in evaluating $S(x)$ for any $x \in [\phi,]$, only four terms at most in the sum (3) will be non-zero.

The basis cubic splines can be constructed by the following recurrent relationship:

$$B_i^n(x) = \frac{x - t_i}{t_{i+n-1} - t_i} B_i^{n-1}(x) + \frac{t_{i+n} - x}{t_{i+n} - t_{i+1}} B_{i+1}^{n-1}(x), \quad (4)$$

$i = -3, \dots, n - 1$ and $n = 1, 2, 3, 4$. A useful convention is to define the first-order splines as *right-continuous* so that

$$B_i^1(x) = \delta_i \text{ rfo} \quad x \in [t_i, t_{i+1}), i = -3, -2, \dots, n + 3, \quad (5)$$

The method is of local character: the change of the position of one control vertex influences only 4 segments of the curve. The resulting curve is in particular coordinates a polynomial of $3 - rd$ degree for $t \in (t_j, t_{j+1})$ and has continuous all derivatives in these coordinates.

Since $B_i^n(x)$ is nonzero only on the interval $[t_i, t_{i+4}]$, the linear system for the B -spline coefficients of the spline to be determined, by interpolation or least-squares approximation, is banded, making the solving of that linear system particularly easy.

$$S^4(x_j) = \sum_{i=0}^n B_i^4(x_j) a_i = y_j, \quad j = 0, \dots, n \quad (6)$$

for the unknown B -spline coefficients a_i in which each equation has at most 4 nonzero entries.

We selected four interpolation methods:

- a) Cubic spline interpolation with Lagrange end-conditions (*cs1*) (i.e. it matches end slopes to the slope of the cubic that matches the first four data at the respective end);
- b) Cubic spline interpolation with not-a-knot end-condition (*cs2*);
- c) Spline interpolation with an acceptable knot sequence (*cs3*);
- d) Spline interpolation with an optimal knot distribution (*cs4*). As 'optimal' knot sequence the optimal recovery theory of Micchelli, Rivlin and Winograd [3] is used for interpolation at data points $\tau(1), \dots, \tau(n)$ by splines of order k ;

All the discussed interpolation methods satisfy the Schoenberg-Whitney conditions, i.e. the achieved representation is for the method, the given data and knot sequences unique. These methods were applied to each of the intervals of the curve, and compared from the point of view of their approximation error (least mean square of the differences of the input value and the spline value) on the given data.

We made a surprising observation: Spline interpolation with an acceptable knot sequence in all intervals of all profiles approximated the data with a smaller error than spline interpolation with optimal knot distribution.

We select an 'optimal' method according to the following criteria: The first criterion for selection of the most appropriate interpolation method is the minimal approximation error on the data in the corresponding interval. To exclude ambiguity, the second criterion is applied: minimal length of the knot sequence corresponding to the method. To further exclude ambiguity, the third criterion is applied: the priority of the interpolation method based on the statistical observations. The priority of the methods was achieved experimentally on profiles and their particular intervals and expresses a 'statistical' ordering according to the smallest approximation error over all intervals of the tested profiles.

2.2 Approximation by Cubic B-Splines

Since in the task being solved, the amount of data pairs acquired to describe a vessel or its parts do not always suf-

fice to represent the shape of the vessel reliably, interpolation does not have to be always the appropriate method. From this reason, we compared the approximation methods on representing the overall shape of the whole curve with respect to the interpolation methods.

The following approximation methods were applied and compared:

- a) Cubic smoothing spline with the smoothing parameter $p > 0$ (*cs5*);
- b) Smoothing spline with the smoothing parameter $t \rightarrow 0$ (*cs6*);
- c) Least squares spline approximation with the number of knots equal to a half of the amount of the data (*cs7*);
- d) Least squares approximation with the number of knots equal to the number of data - degree of the spline in the particular interval, (*cs8*);

3 Results

When 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. This representation is unique and optimal with respect to the above-mentioned criteria. The method was tested on profiles like shown in Figure 4.



Figure 4. Profiles of different fragments.

All interpolation and approximation methods are applied for every sub-interval of the curve after each run of the program. While the curve is generated gradually for each sub-interval of the curve, the overall approximation error is computed. As a result the profile is constructed from the selected methods and is compared to the data set. Table 1 displays the approximation errors for all methods in all intervals of the leftmost profile in Figure 4, including the selected interpolation and approximation methods for the corresponding interval and the selected overall method for the whole profile. The whole data sets contained approximately 350 data points and the length of the whole curve was approximately 400 points.

The most frequently selected interpolation method was *cs1* and the most frequently selected approximation method

was *cs6* in our experiments. An interpolation method was preferred in the intervals, where a sufficient number of data with respect to the length of the interval was given. An approximation method was preferred in the intervals, where there was a lack of data. Figure 5 right half shows one example of an automatically segmented pot with the characteristic points detected and the appropriate manual segmentation on the left of Figure 5 .

method / interv.	1	2	3	4
cs1	0.2163	0	0.6047	0.0781
cs2	0.2163	0	0.5994	0.0782
cs3	0.2163	0	0.5994	0.0782
cs4	0.2163	0.6169	2.1080	0.0877
cs5 ($tol = 5$)	0.2163	2.3114	0.5994	1.1816
cs6 ($p = 1$)	0.1350	0	0.6229	0.07812
cs7	0.2163	5.9470	5.5298	0.5015
cs8	0.2163	0.0032	0.6014	0.1308
select. intp.	1	1	2	1
select. appr.	6	6	5	6
overall select.	6	1	2	1
method / interv.	5	6	7	8
cs1	1.1685	2.2497	1.1424	0.0884
cs2	1.1686	2.2514	0.1433	0.0884
cs3	1.1686	2.2514	0.1430	0.0883
cs4	1.4510	2.3485	0.1615	0.0991
cs5 ($tol = 5$)	2.9430	2.2514	2.2073	0.0884
cs6 ($p = 1$)	1.1687	2.2496	0.1646	0.0884
cs7	6.9127	6.2323	0.8617	1.0675
cs8	1.1850	3.8347	0.1430	0.2551
select. intp.	1	1	1	1
select. appr.	6	6	8	6
overall select.	1	6	1	6

Table 1. Approximation errors for all methods in all intervals.

4 Conclusion and Outlook

The method presented for selection of an 'optimal' representation (optimal with respect to the considered methods and selection criteria) of a 2-dim profile of an archaeological fragment computes and displays a unique solution. The achieved fragment representations, the first part of an automated system for classification of archaeological fragments, are the input of the second part, the classification.

The profile parts, the so-called profile primitives, are used to perform the classification. The segmentation (division) into primitives depends on the orientation of the fragment. In order to achieve a unique representation, it is important to set a unique orientation for all fragments. The classification will be solved in the high dimensional real space and therefore the uniqueness and the high precision of the profile representation are very important.

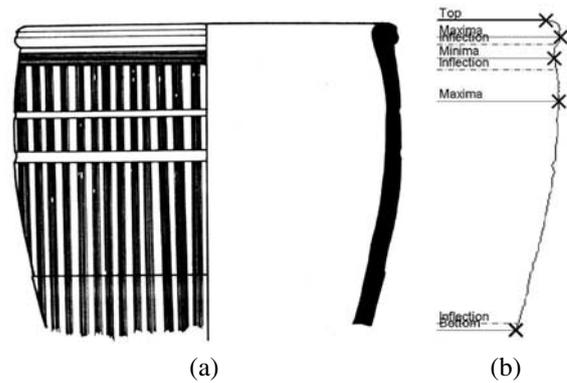


Figure 5. classified pot, (a) manual drawing, (b) detected characteristic points for primitive classification.

The method has been tested on synthetic and real data with good results. The current task is to meet the archaeological requirements as for the achieved representation.

References

- [1] W. Adams and E. Adams. *Archaeological Typology and Practical Reality. A Dialectical Approach to Artifact Classification and Sorting*. Cambridge, 1991.
- [2] J. Bennett and J. MacDonald. On the Measurement of Curvature in a Quantized Environment. *IEEE Trans. Computer*, 24:803–820, 1975.
- [3] T. R. C.A.Micchelli and S. Winograd. The Optimal Recovery of Smooth Functions. *Numerische Mathematik*, 26:191–200, 1976.
- [4] R. DeVore and G. Lorentz. *Constructive Approximation*. Springer, 1993.
- [5] Z. Hu and S. Ma. The Three Conditions of a Good Line Parameterization. *PRL*, 16:385–388, 1995.
- [6] M. Kampel and R. Sablatnig. On 3d Modelling of Archaeological Sherds. In *Proceedings of International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, Santorini, Greece*, pages 95–98, 1999.
- [7] C. Liska and R. Sablatnig. Estimating the Next Sensor Position based on Surface Characteristics. In *15th. ICPR*, pages 538–541, September 2000.
- [8] J. Matas, Z. Shao, and J. Kittler. Estimation of Curvature and Tangent Direction by Median Filtered Differencing. In *8th International Conference on Image Analysis and Processing*, pages 83–88, 1995.
- [9] C. Orton, P. Tyers, and A. Vince. *Pottery in Archaeology*, 1993.
- [10] P. Rice. *Pottery Analysis: A Sourcebook*, 1987.
- [11] A. Rosenfeld and A. Nakamura. Local Deformations of Digital Curves. *PRL*, 18(7):613–620, July 1997.
- [12] A. Shoukry and A. Amin. Topological and Statistical Analysis of Line Drawings. *PRL*, 1:365–374, 1983.
- [13] C. Sinopoli. *Approaches to Archaeological Ceramics*. New York, 1991.

Chapter 5

Structural Analysis of Paintings Based on Brush Strokes (“Art History”)

Robert Sablatnig and Paul Kammerer and Ernestine Zolda. Structural Analysis of Paintings Based on Brush Strokes. In *Proc. of SPIE Scientific Detection of Fakery in Art*, volume SPIE-Vol.3315, pages 87–98. San Jose, USA, January 1998. SPIE (invited paper).

also in:

In W. C. McCrone and R. J. Weiss, editors, *Fakebusters Scientific Detection of Fakery in Art*, pages 222–244. Hansen, Stoughton, Massachusetts, 1999.

Structural analysis of paintings based on brush strokes

Robert Sablatnig, Paul Kammerer, Ernestine Zolda

Pattern Recognition and Image Processing Group

Vienna University of Technology, Treitlstr. 3, 183/2, A-1040 Vienna, AUSTRIA

Fax: ++43 /1/ 505 4668, e-mail: sab@prip.tuwien.ac.at

ABSTRACT

The origin of works of art can often not be attributed to a certain artist. Likewise it is difficult to say whether paintings or drawings are originals or forgeries. In various fields of art new technical methods are used to examine the age, the state of preservation and the origin of the materials used. For the examination of paintings, radiological methods like X-ray and infra-red diagnosis, digital radiography, computer-tomography, etc. and color analyzes are employed to authenticate art. But all these methods do not relate certain characteristics in art work to a specific artist - the artist's personal style.

In order to study this personal style of a painter, experts in art history and image processing try to examine the "structural signature" based on brush strokes within paintings, in particular in portrait miniatures. A computer-aided classification and recognition system for portrait miniatures is developed, which enables a semi-automatic classification and forgery detection based on content, color, and brush strokes. A hierarchically structured classification scheme is introduced which separates the classification into three different levels of information: color, shape of region, and structure of brush strokes.

Keywords: classification, art history, brush strokes, feature detection, face detection, segmentation.

1. INTRODUCTION

The Austrian royal family members remaining in Vienna were connected with their relatives in foreign countries through a collection of nearly 600 *portrait miniatures* (small format pictures of approximately 8cm x 6cm of a person which basically has the same character as a photograph, see Figure 1), started by Empress Maria Theresia. The artistic manifestations in the collection cannot lead to a clear affiliation with certain artists, especially because of the lack of signatures on the portrait miniatures.

Although it is undeniable that an artist produces paintings of varying quality, and that these variations can be considered as a whole by an art historian the question still remains whether or not an artist can be identified by certain replicable constants, and how these are described [7,13]. In the case of portrait miniatures painted in an aquarell style using point and line strokes, one can recognize certain mechanical trademarks. Each artist applies different hues to the strokes or points to create a distinct portrayal of the subject. Because the subject is produced in a very small format and the application of multiple strokes is used to create the face, one can observe that the artist has relied on his/her own unconscious *rhythm*. This term describes not only the "handwriting" of an individual artist which follows a certain pattern of stroke length and angle, but also the system of lines and the relation of lines to one another. However, personal patterns can be recognized, in both high quality and mediocre paintings of an artist. In the case of portrait miniatures [2], this basic pattern is that of the face, which is constituted of different recurrent parts (basic oval form, eyes, nose, mouth, etc.). The artist applies an individual line system of strokes and colors to this basic



Figure 1: Portrait miniature

pattern, thereby not only adding his artistic signature but also giving the portrait an individual physiognomy.

To meet this identification problem we develop artist specific models, which describe artist specific and artist independent characteristics of a painted portrait miniature. We claim that these characteristics are expressed in the way the artist placed the brush strokes and the constraints he has to consider to create a realistic reproduction of a human face [27]. The overall-goal of our work is to support the art his torian expert in order to classify and identify the miniatures in an objective manner.

The method developed so far concentrates on the detection of stroke segments in intensity images of the portrait minia- tures. The following section gives a short overview on the image acquisition process and the classification scheme that will be used to relate miniatures to artists. In Section 3 a description of the classification model, where artist specific features are represented is shown, the local stroke-detector is presented in Section 4 in detail. An analysis of the results of the stroke- detector in Section 5 discusses the applicability of the proposed classification scheme. The paper concludes with an outlook on further working steps.

2. IMAGE ACQUISITION AND CLASSIFICATION SCHEME

To perform a semi-automatic classification of the portrait miniatures, they were digitized using a 3-chip color CCD camera. This color image contains only the face of the person since this part contains the artist relevant information in contrast to the background. This color image of the face (size: 572x768 pixels) is the basis for the subsequent classification.

The structural analysis to relate an artwork to an artist can be divided into 4 hierarchical classification steps: *Color classification*, *Shape classification*, *Stroke classification*, and finally the *Artist classification*. This global top-down classification scheme is extended by a bottom-up strategy within each classification step. Figure 2 shows the single steps of the clas- sification scheme. The process starts with the *Image acquisition*, which is the basis for the color classification. Next, a *Color space transformation* is used to reduce the amount of data and to perform the subsequent intensity based feature detection. An intensity image is the basis for the *Face extraction*, the face is segmented from the background. Within the face classification relevant **Regions Of Interests (ROI's)** are produced with the help of *Region Segmentation*. These ROI's are used to compare shape features (like eyes, mouth, and nose) within these ROI's to shape features of different artists for shape classification. Within certain ROI's *Stroke detection* is performed to capture the "rhythm" of the artist. Since this detection computes only stroke segments,

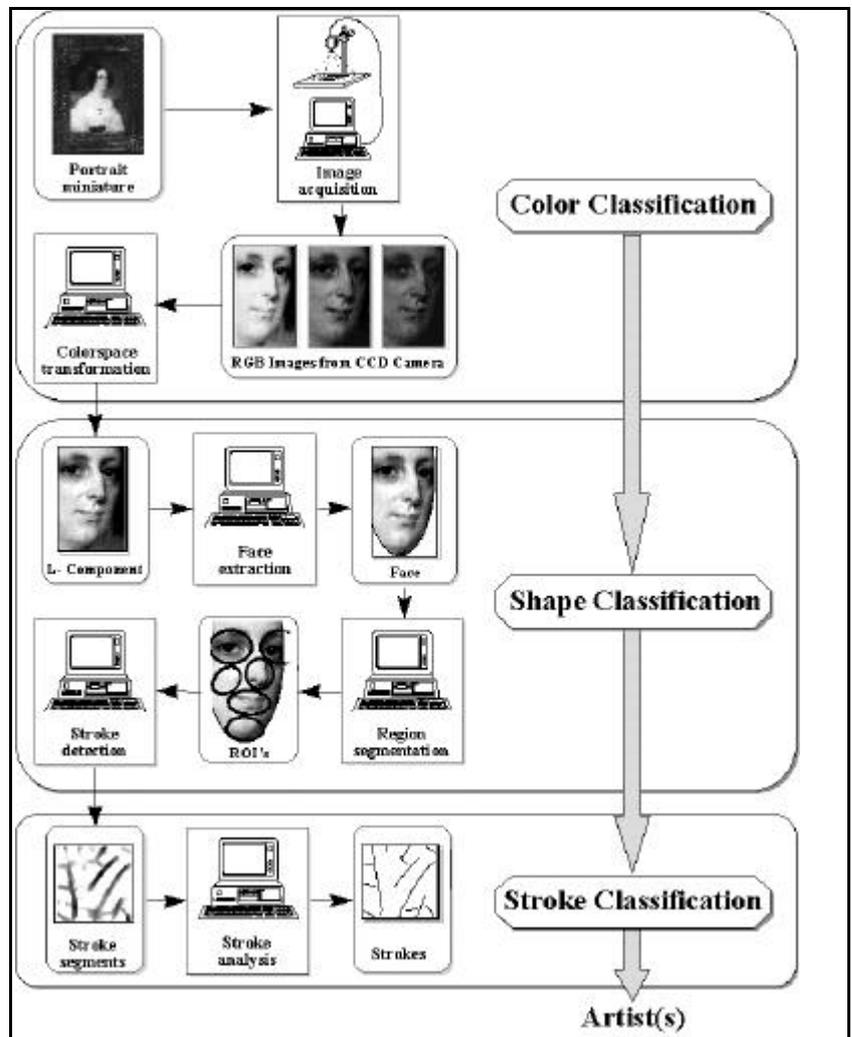


Figure 2: Image acquisition and classification scheme

a subsequent *Stroke analysis* is necessary to assemble these segments into strokes. The stroke classification within certain regions allows a robust artist classification. In the following a brief description of the function of the individual processing steps is given:

- **Image acquisition:** To meet the art-historical requirements, the acquisition system is able to register the face details in true color (24-bit). This is attained by the use of a 3-chip color CCD-camera in combination with a specific optical lens system. A true-color representation of the image is necessary to carry out the color classification. The signals produced by the camera are digitized by a color framegrabber device and prepared for further processing steps. During the acquisition process the illumination conditions were held constant.
- **Color classification:** Artist classification by art historians is based on color impression. This term describes the overall color perception of the painted face - it's color tone. This color impression considers the origin of the object, because colors and the manner of using them vary with time. Based on this fact, portrait miniatures are first grouped by color. To perform this rough classification automatically, the mean RGB value of the face is calculated. But RGB color values represent only one part of the color impression of an art historian, he judges also the person presented (man or woman), the age of the showed person, the painting manner, and the orientation of the face. Nevertheless, this color classification is used to perform a rough pre-classification to separate miniatures because of their color tone.
- **Color space transformation:** Color space transformations [21] are used as a preprocessing step to prepare the registered images for the extraction of the features. Since true color image processing is a time-consuming process and there is a lack of feasible methods of color feature detection [4,14], we decided to develop a stroke detection method based on intensity images. Hence it is necessary to transform the image data from the color space to a gray value space without losing too much information with regard to stroke detection [12]. We therefore chose a RGB to HLS transformation, because the L (lightness) channel of the HLS-model represents the information (intensity values of the color image), which is relevant for line detection [26].
- **Face extraction:** Since relevant artist specific information can only be found within the face of the shown person it has to be segmented from the background. This segmentation is done for two reasons; to reduce the search space and hence the computation time and to be able to segment the face into certain regions like eyes, nose and mouth to name the most prominent. This segmentation can be performed easier than conventional face detection since these faces are created by an artist which has a standard "creation model" in mind (in contrast to real perspective projections). Artists use a scheme which they have learned during their practice to draw faces. Especially in portrait miniatures the painting style is more schematic than realistic. There is evidence [7] that artists used an elliptical schema, which determines the shape and orientation of the head as well as the position of eyes, nose, cheeks etc., to develop their portrait miniatures (see Figure 3a).

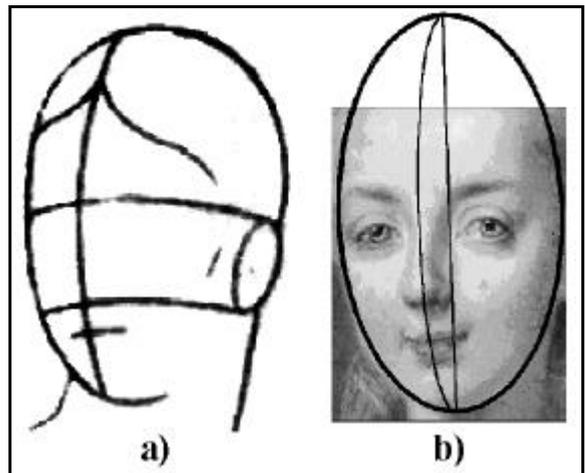


Figure 3: a) schematic head shape [7], b) fitted ellipses

Our approach assumes that the head can be described by an ellipsoid whose major axis is parallel to the image plane. The projection of the ellipsoid into the image plane is an ellipse (*face ellipse*), which is invariant to the rotation of the ellipsoid with respect to the major axis. A second ellipse within the ellipsoid describes the intersection of the facial symmetry plane with the ellipsoid (*symmetry ellipse*, i.e. the ellipse which lies in the direction of the nose). The ellipses are up to now adapted manually to the face contours by setting the focal points and a point lying on the ellipse [18]. The second ellipse shares the major axis of the first ellipse and its minor axis is oriented according to the orientation of the face. Figure 3b shows the result of the ellipse fitting, the face ellipse, the symmetry ellipse and the rotation axis of the rotational ellipsoid.

- **Region segmentation:** Following the face detection and ellipse fitting the image plane ellipse is segmented into different facial regions because the stroke arrangement differs within certain regions. The stroke arrangement differs for instance significantly between eye and the cheeks regions due to their different shape and illumination. To estimate the position of the head, to segment the face into characteristic regions, and to support the extraction of facial features, two ellipses (face ellipse and symmetry ellipse) are used. The information about the rotation of the face ellipse is used to determine elliptic regions like eye, nose, mouth, chin, and cheek [16]. Figure 4 shows the segmented regions within the face.
- **Shape classification:** The face segmentation allows a comparison of portraits on a region by region basis. Artists tend to use a schematic way to model face details. For instance eyes within portrait miniatures are painted rather schematic than realistic, the depicted eye is only similar to the eye of the painted person. This fact is used to classify miniatures due to the shape of certain regions, some artist tend to paint "circular" eye regions, some tend to paint them in elliptic shape to give an example. The region based matching reduces the complexity of comparison too since the search space is reduced from the complete face to specific ROI's.
- **Stroke detection and analysis:** In order to compare the segmented regions not only by their shape but also by the brush strokes used to paint them, a segmentation of the individual strokes is necessary. One way to segment the strokes is to apply edge detection methods that are widely known in pattern recognition. After having studied the results of standard edge detection methods, we found that stroke detection is only practicable if an accurate model of a brush stroke is defined beforehand. The results of a stroke detection method based on a model incorporating parameters like length, width, curvature and so on are given by images of detected stroke segments in various orientations. These brush stroke segments are grouped into brush strokes by matching similar curvatures and orientations of neighboring stroke segments. If they match within certain tolerances, these stroke segments are connected forming the original brush stroke.
- **Stroke classification:** The detected strokes are needed to carry out a structural analysis with regard to a classification. The structure of the detected stroke segments allows a classification of the miniatures since the brush stroke structure is very similar to the basic elements of art historical classification. Furthermore it can be recognized that there are similar arrangements in all miniatures (e.g. contour lines). Therefore, there are two different kinds of characteristics: artist dependent characteristics given by the individual "handwriting" of the artist and artist independent characteristics as a constraint from assumed illumination, shape, and 3D-impression of the face.

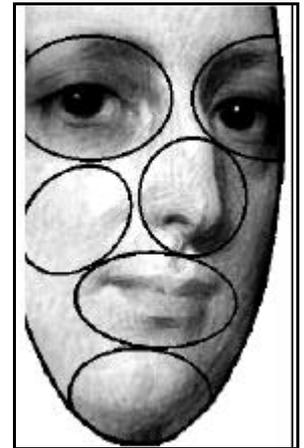


Figure 4: Segmented ROI's

The hierarchical structure of the three classification steps allows a top-down classification, first the color impression is used to have a rough classification, within this subset of all possible artists a more detailed classification is performed due to the shape features within certain face regions which reduces again the set of possible candidates. Within the regions of the face under examination the stroke classification introduces a bottom-up approach. This three classification stages allow an economic search space adaptation, having a set of possible artist as classification result. The art historian judges this result, if it is not correct, the artist model has to be adjusted.

3. PORTRAIT CLASSIFICATION MODEL

To relate the characteristics to a specific artist, a so-called *artist-model* is developed. The model is based on a reference model of a human face, which assigns local artist dependent characteristics to facial regions. A closer look at portrait miniatures and their corresponding painting technique reveals that there are a number of brush strokes and brush stroke arrangements that are not only specific for a certain artist but also influenced by the shape of the painting ground and the object to be painted. To achieve a realistic impression of a face - the portrait of the person painted should be recognizable - the artist must consider the influence of physical illumination on the appearance of the face. That means the artist has to consider the direction of the incoming light and the 3D shape of the face for each individual point of the face to be illustrated. Therefore, we have to divide the model parameters into parameters determined by the specific illumination situation and parameters that describe the individuality of the artist. To answer the question: "is a certain local arrangement of strokes due to the individuality of the artist or due to the working process?" we take a closer look at the 2d- brush strokes. To achieve this, a mathematical reference model of a human face that is connected with the image information is introduced. In addition to the image information this model provides geometric information to support the separation into artist-specific and artist-independent components structured into specific regions.

A connection between the three-dimensional world coordinate system $(x,y,z) \in \mathbb{R}^3$ of a human face and a parameter space $(s,t) \in \mathbb{S}^2 \times \mathbb{T}$ can be defined as follows (see Figure 5): N : nose = $(0,0)$; M : mouth = $(0,-1)$; A_1 : left eye = $(-1,2)$; A_2 : right eye = $(1,2)$. The geometrical connection between arbitrary points in the parameter space (s,t) and the corresponding points in the world coordinate system can be defined via a homomorph transformation: $\mathbf{J}(s,t) = (x(s,t), y(s,t), z(s,t))$.

Each point in the parameter space is assigned to radiometric information. If we use for instance the HLS color system, we have three information components, Hue $h(s,t)$, Lightness $l(s,t)$, and Saturation $s(s,t)$. Brush strokes can be represented in the parameter space by ellipses with a certain angle $q(s,t)$, which defines the orientation of the stroke, and with a certain length of the primary axis, which defines the length of the stroke. In addition, representations of elongated regions ("ribbons") [25] can be applied. Cross-hatched and parallel strokes can be represented as a combination of individual strokes.

This reference model is based on knowledge from the field of face recognition [10,11,23], which also uses reference models to represent a human face and which can be transformed if needed. The definition of the individual reference points and distances has its origin also in this field. Another research area our model is connected with is "Shape from Shading" [9,22].

This mathematical model has many benefits. For example, we can select a certain elliptic region within the parameter space of the reference model and transform it to the appropriate image via geometric connection. In this way we can compare a cheek region from one artist with the cheek region from another artist, although the one artist might have used a frontal and the other a side view. Figure 6 illustrates the application of the mathematical model to a miniature.

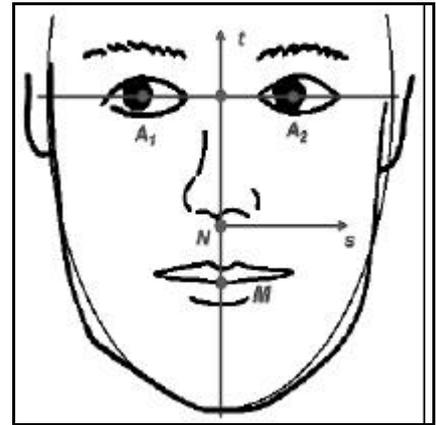


Figure 5: Mathematical model of the face

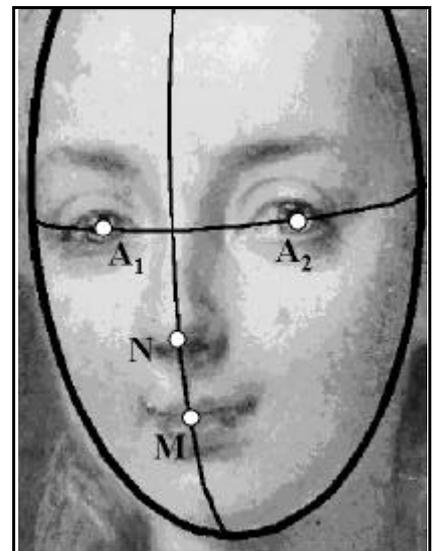


Figure 6: Reference points of the model

Further restrictions concerning the detection process are given by the implicit goal of an artist to produce an image of the person which is supposed to be represented as realistically as possible. This means that he has to consider the position of the light source as well as the shape of the actual face and consequently use a certain lightness shading. In other words, in certain areas of the human face strokes have to be placed primarily in certain directions in order to achieve a sense of plasticity. This fact is illustrated in Figure 7 as brush strokes of a portrait miniature are traced by hand.

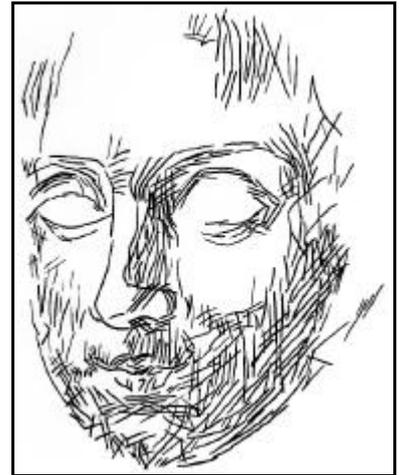


Figure 7: Brushstrokes traced by hand

The mathematical reference model is the basis for the artist-specific model. We define artist-specific arrangements of brush strokes, which may appear differently in different face regions. It is important to ascertain whether the arrangement of the strokes reflects the individuality of the artist and is not the result of restrictions by the shape of the face, the projection geometry and so on. Artist characteristic features are represented in a set of parameters (like set of colors, eye shape, face shape average stroke length, width, and curvature to name a few) of the mathematical models. So every artist is assigned to a set of characteristic parameters.

Portrait miniatures which could not yet be classified will be analyzed on the basis of artist-specific models. To compare different artists we need a similarity measure of the artist-specific parameter sets. The verification of the model within an image results in a measure, which allows to distinguish paintings of artists.

4. STROKE-DETECTOR

The final detail classification is based on the arrangement of strokes. Therefore, the stroke detection is the most significant step in order to perform the classification. Finding brush strokes in an image of portrait miniatures is a typical segmentation problem. The aim is to separate the strokes from the background. Two depicted regions are distinguishable since they differ in color and/or brightness. Since our focus is on the facial detail of portrait miniatures, color difference is not relevant in our specific case, and thus we concentrate on brightness differences instead [15].

The first attempt to such kind of detection problems is the application of standard edge detection methods to the lightness components of four test images and judging the attained results. To demonstrate and discuss the results, a detail (nose detail Figure 8a) of a portrait miniature is used. Figure 8b shows the detail superimposed by markers of significant strokes determined by the art historian. The goal of edge detection is to find the borders of the brush-strokes. A non-maxima suppression method [24] was applied to the gradient images to evaluate the results. The result is a binary image in which each pixel represents a locally maximal gradient. The resulting binary image of the first edge detector applied, the Sobel edge detector, can be seen in Figure 8a. The image shows the difficulties of segmentation using a global threshold. The result of the Canny edge detector [3] applied to the image of Figure 8a can be seen in Figure 8b. To produce the binary image, we used a threshold calculated according to Canny's hysteresis method using values of 1.5 and 0.75. The last detector of our experiments is known as Marr-Hildreth or DoG (Difference of Gaussians) detector [20]. Figure 8c shows the resulting image using $\sigma = 1=9$.

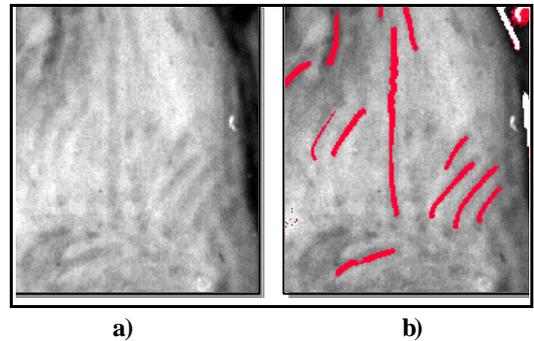


Figure 8: Test image: a) intensity image, b) significant brushstrokes

Using the three resulting images of Figure 8 we can compare and interpret the application of standard edge detectors, like Sobel, Canny, and Marr-Hildreth edge detector [1,6], in view of our specific problem. We sought to detect the borders of the individual strokes, which correspond to edges in an intensity image. It is known that the size of the filter kernel influences the detection result [8]. The result image of the Sobel operator shows the considerable response to local intensity changes due

to a kernel size of 3 by 3 pixel. Both the Canny and the Marr-Hildreth detector use a larger kernel resulting in a smoothing effect over a larger region. This leads to a weaker response to small intensity changes in the image. It can be pointed out that further postprocessing steps are necessary, if we want to use the detected edges as a representation of stroke borders. The determination of strokes based on edge detection is expensive due to broken edges. Therefore, more knowledge about the origin of brush strokes into the detection process is integrated. One possibility is to use a line detection method instead of an edge detector. The method that will be presented here is based on a stroke model which describes brush strokes as line-like objects.

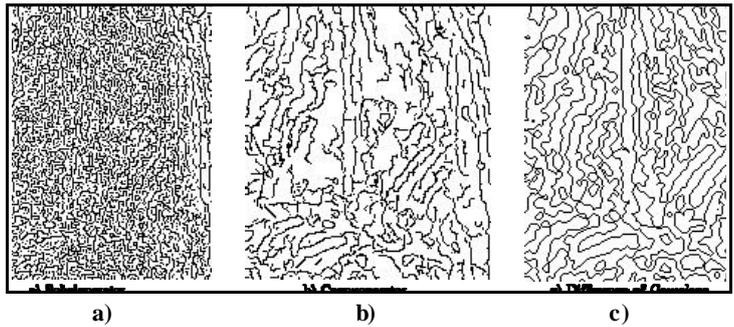


Figure 9: Results of standard edge detectors: a) Sobel b) Canny c) DoG

The method that will be presented here is based on a stroke model which describes brush strokes as line-like objects.

4.1. Brush stroke model

This section deals with fundamental questions, like how to characterize a brush stroke, what kind of information is necessary for a viewer to perceive a brush stroke within the detail of a face and how can a brush stroke be reliably identified by a human viewer. Furthermore we will discuss the different views of technicians and art historians, respectively and how both views can be reconciled into one. The following brief "ad hoc" definition of the term "brush stroke" illustrates the different view points:

- ! Both art historians participating in the project defined a brush stroke as follows: A brush stroke is a stroke painted by an artist. It can be distinguished from the background and thus be perceived by the human eye. A brush stroke contains a point of origin and an ending point, which is darker and more saturated due to the lifting of the brush.
- ! The following definition concerning the appearance of a brush stroke within a digital image of a portrait miniature was given by the project participants with technical background: a brush stroke is an elongated shaped region of pixels which are characterized by individual color information.

The statement from the art historian point of view of that a brush stroke can be recognized only if it stands out against the background can be generalized in connection with the faculty of human perception. The properties of a brush stroke gathered so far can be summed up as follows:

- ! brush strokes are painted using a fine brush;
- ! some strokes painted in a wet, and others in a dry manner;
- ! brush strokes show some colors;
- ! brush strokes are elongated objects;
- ! depending on the brush, brush strokes have a certain length and a width;
- ! strokes show a direction, which means that there is a point of origin and an end point;
- ! there are combinations of strokes such as: parallel strokes, cross hatched strokes, etc.;

Based on this facts an experiment was made to answer the following two essential questions:

1. How reliably can a brush stroke be recognized by the human eye?
2. What size of background of a stroke is necessary in order for a person to be able to recognize the stroke?

We used four test images of portrait miniatures in the experiment. The art historian marked significant strokes in the image by hand. 120 optically significant strokes, distributed over the whole image, were selected. The selected strokes were presented to the art historians and to the technicians respectively in a various sized environment. The task was to decide whether a stroke can be identified as such or not. The mean recognition rate of all participants was 84.2 %. This value was

obtained from a neighborhood size three times the width of a brush stroke. The results of our experiment permit us to make the following assertions concerning the definition of a brush stroke model and the development of a stroke-detector:

- ! the stroke can be recognized in a locally;
- ! the neighborhood must have a certain minimal size, which is approximately three times the width of the stroke;
- ! strokes are darker than the background, since in the specific art category of portrait miniatures bright colors are used first, dark colors are painted afterwards and therefore the last layer of strokes is locally darker than the background;
- ! strokes have different orientations; have a certain curvature; and may be interrupted by other strokes.

Using this knowledge of brush strokes in portrait miniatures we can define a model of a brush stroke which is based on the medium axis of the stroke. Figure 10 gives an illustration of the model and its parameters.

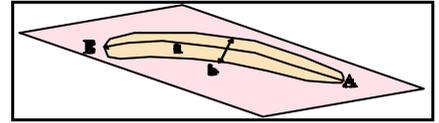


Figure 10: Parameters of a brush stroke

The medium axis of the brush stroke is represented by a parametric curve $a(t)$. The curvature $k(t)$ of the brush stroke as well as the orientation of the stroke $\alpha(t)$ can be computed from this axis. Point A and point E correspond to the point of origin and the end point of the brush stroke. The path from A to E represents the painting direction of the brush stroke.

! axis of the brush stroke: $a(t) = (x(t), y(t)), t \in [0,1]$

! starting point: $A = a(0)$

! end point: $E = a(1)$

! width at position t : $b(t)$

! length:
$$l = \int_A^E \sqrt{\dot{x}^2(t) + \dot{y}^2(t)} dt$$

! curvature at position t :
$$k(t) = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}$$

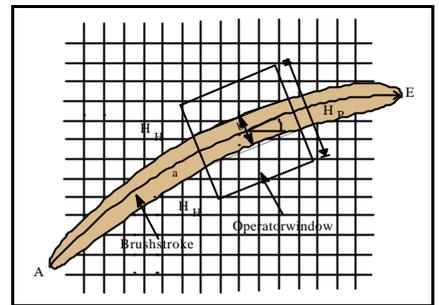


Figure 11: Model of a brushstroke

The model defined above can be used to determine the features of a brush stroke. Furthermore the following properties of a brush stroke will be used to create a local stroke-detector.

1. Stroke detection can be performed locally
2. The size of the detector kernel is about three times the stroke width
3. Generally the region corresponding to the strokes is darker than the background
4. The detector must be able to detect strokes in every position and orientation within the image

These properties can be used to define a stroke-detector based on a line model of a stroke. Figure 11 illustrates the model and the parameters of the model where b depicts the width of the brush stroke, $3b$ the size of the detector kernel, HP the brightness of the brush stroke, HH the brightness of the background ($HH > HP$), and $\hat{\alpha}$ the local orientation of the brush stroke. A detector to detect brush strokes was developed on the basis of this model.

4.2. Detecting brush strokes

The brush stroke model provides the information for the choice of a detector. Table 1 shows the connection between the parameters of the model and the corresponding detector specifications.

A local line detector able to detect lines in various orientations fulfills the requirements of the model. Since linear line detectors also detect edges, our stroke-detector is based on the semilinear "streak" detector described by Rosenfeld [24]. We applied modifications to make it possible to detect strokes in any direction while introducing a nonlinear condition which helps to avoid false detections.

Model parameters	Detector
location: brush strokes can be recognized locally	local detector
minimal size of a stroke environment for stroke recognition	size of the detector kernel
brush strokes are darker than the background	corresponds to positive contrast line model
brush stroke can appear in arbitrary orientations	detector for strokes of any orientation

Table 1: Model parameters influencing the detector

When using a line detector the number of possible orientations is in the range of 0 to 180 degrees since the direction of a line is not taken into consideration, in contrast to an edge [5]. The detector provides an intensity image representing the line gradients for each of the orientations used. The images of the individual gradients are combined by forming the maximum of the images. Let G_i be the gradient image of orientations $O(i)$ with $i = 1..N$ and N be the number of orientations at an angular distance of $N/180$. The combination image (maximum image) M is calculated by computing the maximum for every pixel at position (x,y) for all gradient images G_i : $M(x,y) = \text{MAX}(G_i(x, y))$.

The detector was applied to the image in four different orientations of equal angular distance (0E,45E,90E,135E). The result of the detection, four intensity images, represents the response of the detection. The gray values may be interpreted as the strength of the line gradients. Strokes that are rich in contrast in the intensity image show high values in the gradient image. However, it should be pointed out that the gradient strength depends on the local contrast of the image. On the uniform background, a very dark thin line may result in the same gradient strength as a less darker wide line. The detector provides a line gradient image for any orientation it is applied to. The gray values of the gradient image can be seen as a measure of the brush stroke's contrast with the restrictions mentioned above. Figure 11c shows the maximum image of eight gradient images which are the result of applying the line detector to the intensity image (Figure 11a) in eight combined orientations.

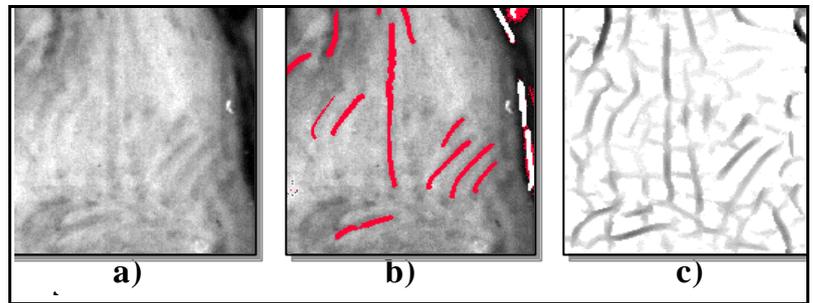


Figure 12: Results of the application of the stroke detector: a) intensity image, b) manually determined strokes, c) result of stroke detection (inverted)

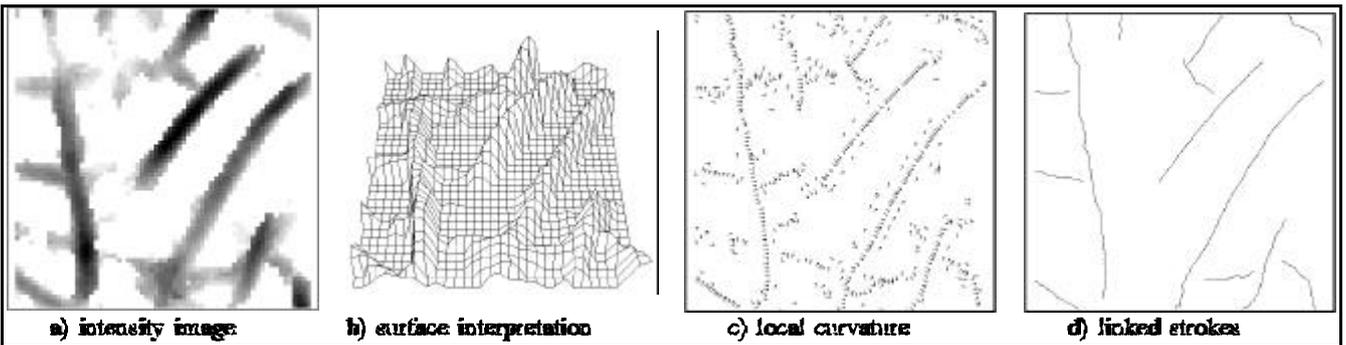


Figure 13: a) nose image, b) intensity surface interpretation, c) points of maximal intensity with directions, d) result of linking

The maximum image consists of parts of brush strokes so-called stroke segments. The crossing and overlapping of strokes cause incorrect results of the detector. Therefore, these stroke segments must be linked to form complete strokes, which are interpreted by the human viewer as a consequence of interpolation. In addition parameters like length, width, curvature and location are calculated for the individual stroke, which is used as input for the classification. Furthermore it must be anticipated that some strokes will be covered by other strokes. To assign the stroke segments to original strokes we

assume the output of the detector as a surface in the 3D space where the intensity value becomes the surface height. Figure 13a shows a detail of Figure 11c; Figure 13b shows the intensity surface interpretation of the detail on which the detected stroke segments appear as "ridges". We compute the location of points with high positive transverse curvature and locally maximal height in the transverse direction in subpixel accuracy. The computed points with the corresponding curve directions are shown in Figure 13c, Figure 13d shows the results of linking the computed points, based on the notion of a Euclidean distance neighborhood. Most of the strokes (80%) are detected correctly, but in some cases it is necessary to use the curvature, the direction and the color of strokes as additional parameters to improve the result of the linking process.

5. STROKE-DETECTOR ANALYSIS

In a test series we applied the stroke-detector to 44 intensity images of miniatures using 8 distinct orientations with equal angular distance (0,22.5,...,167.5 degree). To verify the results an art historian segmented the brush strokes manually (by marking the strokes with a pen on the monitor into an overlay) to prepare a *ground truth* (this segmentation is of course subjective since only strokes with high contrast are marked but it this is a first approximation of the human perception of strokes). Beside regions, where individual strokes can be identified by the human observer, there are facial regions that show only few recognizable strokes.

Subimages (all placed in the forehead of the face) were used to analyze the detector results in relation to the detector window size, i.e. how detection rate of strokes depends on the detector size. According to the brush stroke model the stroke width of 3, 5, and 7 pixel corresponds to a detector window size of 9x9, 15x15, and 21x21 pixel respectively. To give an example, Figure 14a shows a detail image of a forehead region and Figure 14b four manually segmented strokes. These four strokes are strokes that are immediately separated by the human perception, since they have no interconnection and overlap with other strokes. There are no further recognizable structures within this region, overlapping strokes of low contrast form an almost uniform area. Figure 14c shows the result of the 9x9 stroke-detector (white regions depict strokes) and Figure 14d the result of the 15x15 detector. Strokes shown in Figure 14e were detected by a 21x21 stroke-detector.

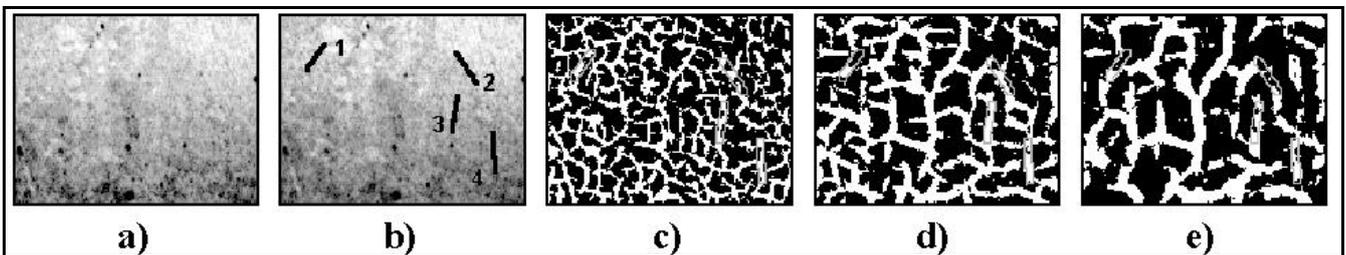


Figure 14: Results: a) intensity image, b) manually segmented strokes, c) 9x9 window, d) 15x15 window, e) 21x21 window

Window size	Stroke 1	Stroke 2	Stroke 3	Stroke 4
9x9 pixels	60%	47%	70%	79%
15x15 pixels	40%	60%	96%	86%
21x21 pixels	42%	25%	71%	78%

Table 2: correspondence between detected strokes and manually segmented strokes (number of overlapping pixels)

The evaluation of the result is based on the correspondence of the manually segmented to the automatically detected strokes, by calculating the percentage of detected pixel within the manual segmented region. The percentage of detected stroke pixels was calculated at detector window size 15x15 pixels (see Table 2). This detector size provides the highest detec-

tion rate because the width of the manually segmented strokes is approximately 5 pixels. Furthermore there is a low percentage of detected stroke pixels at stroke 1 and 2, which are orientated in 45° and 135° . The lower detection rate is due to rotational variances of the square shaped detector window at 45° and 135° , which can only be reduced by a circular shaped detector [12].

The highest recognition rate of manually segmented strokes and the highest number of detected stroke pixels show that there are strokes of width 5 pixel within the almost uniform, bright area of the forehead in this example. In order to identify strokes in the result images using different window sizes, one has to pay attention to the fact, that stroke segments identified by the large detector (21x21) have their correspondence in the result of the smaller detector windows [19]. The artifacts introduced by smaller windows are mainly located at the borders of the stroke segments detected by the larger window size detector. On the other hand we notice also that strokes marked by hand are detected partially by the detector in all different window sizes.

6. CONCLUSION AND OUTLOOK

In this paper a hierarchical top-down classification scheme that uses the color impression as a first rough classification, which is refined in a subsequent shape classification and finally refined in a stroke classification was presented. The stroke classification within certain face regions uses a bottom-up classification scheme, individual strokes are detected and grouped into patterns. To analyze the arrangement of strokes applied by the artist, strokes are detected by a local operator. For the design of the operator expert knowledge was captured in experiments resulting in the definition of a stroke model and its parameters. A local operator, based on a semilinear streak detector [24], was developed according to the stroke model. The parameters of the stroke-detector are the size and the orientation of the detector window which are related with the width and orientation of the stroke. The results of developed stroke-detector were evaluated, it turned out that the detector performs best if the average stroke width is known.

An analysis of the automatically detected stroke results showed, that there are regions in the face which show a lack of manually segmented strokes but a number of stroke segments detected by the stroke-detector. These regions are described by the human observer as the brightest in the miniature, located in the forehead and below the eyes of the unshaded side of the face. Using histogram normalization to enhance the image contrast, some strokes that were not recognized by the human expert were visible afterwards. The comparison between the manual segmented strokes and the detected strokes showed a divergence. The human eye has the tendency to interpolate between partially interrupted strokes and between dark dots resulting from overlapping strokes. Especially in bright regions the human eye cannot decide whether there is a stroke or not. The local stroke-detector in contrast provides a result that is independent from the local contrast.

The experiments made up to now are pre-studies for the stroke classification within the artist specific model, which can describe an individual artist or a group characterized by specific arrangements of strokes. Our goal to classify specific artists within a group of artists is based on a stable method to detect the brush strokes since this feature is the only one that allows an objective distinction between artists - the personal signature. Regions which lack easy recognizable strokes are examined in respect to detector artifacts because of the discrepancy of the number of stroke segments detected by the stroke-detector to strokes identified by the expert. The results of the examination of subimages of the forehead show the stability of the developed method towards different detector window sizes. Future experiments will include further regions which differ in the human perception of strokes and results of the stroke-detector, e.g. the regions below the eyes.

Furthermore we currently work on the region extraction based on the mathematical model and the ellipse fitting. Within these regions shape detection is performed and matched with the shape of the features of different artists. The final goal is the reduction of the number of possible candidate artists to a high degree. The result can facilitate the work of the art historians making the final decision regarding the authorship of a painting. Thus our priority is not the full automation of the process but to support the work of the art historians.

7. ACKNOWLEDGMENTS

This work was supported by the Austrian "Fonds zur Foerderung der wissenschaftlichen Forschung" under grant P12028-MAT. The authors want to thank Walter Kropatsch, Robert Keil and Ingeborg Tastl for the stimulating discussions in the development of this work.

8. REFERENCES

- [1] I. Abdou, W.K. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors", *Proc. of IEEE*, Vol.67, No.5, pp.753-763, 1979.
- [2] L. Becker-Emmerling, "*Die Technik der Miniaturmalerei auf Elfenbein und Pergament*", Ravensburg.
- [3] J. Canny, "A computational approach to edge detection", *IEEE Trans. on PAMI*, Vol.8, No.6, pp.679-698, 1986.
- [4] A. Cumani, "Edge detection in multispectral images", *Comp. Vision Graph. and Image Proc.*, Vol.53, No.1, pp.40-51, 1991.
- [5] P.E. Danielsson, O. Seger, "Rotation invariance in gradient and higher order derivate detectors", *Comp. Graph. and Image Proc.*, Vol.49, pp.198-221, 1990.
- [6] E.R. Davis, "Circularity - a new principle underlying the design of accurate edge orientation operators", *Image and Vision Computing*, Vol.2, pp.134-142, 1984.
- [7] E. Gombrich, "*Kunst und Illusion. Zur Psychologie der bildlichen Darstellung*", Kapitel Erlerntes und Erlebtes, pp.173-205, 1978.
- [8] R.M. Haralick, L.G. Shapiro, "*Computer and Robot Vision*", Vol.I&II, Addison-Wesley, 1992.
- [9] B.K.P. Horn, "Understanding image intensities", *Artificial Intelligence*, Vol.8, pp.201-231, 1977.
- [10] K. Hrabý, "*Identifizieren von Gesichtern durch Steuerung der visuellen Aufmerksamkeit*", Master's thesis, Vienna Univ. of Technology, 1996.
- [11] M.S. Kamel, H.C. Shen, A.K.C. Wang, T.M. Hong, R.I. Campeanu, "Face recognition using perspective invariant features", *Pattern Recognition Letters*, Vol.15, pp.877-883, 1994.
- [12] P. Kammerer, E. Zolda, "Pre-studies for artist specific models for the pre-classification of portrait miniatures", in: *Proc. of OAGM97: 21st Workshop of the Austrian Association for Pattern Recognition*, Hallstatt, Austria, May 26-27, 1997, in press.
- [13] R. Keil, "*Proportionslehre und Kunstbuchliteratur als künstlerisches Lehrmaterial im 16. Jahrhundert - Versuch einer Entwicklung nach Dürer unter besonderer Berücksichtigung der italienischen Tradition*", PhD thesis, University of Vienna, 1983.
- [14] A. Koschan, K. Schluess, "*Grundlagen und Voraussetzungen für die digitale Farbbildverarbeitung*", Technical Report 94-14, Fachbereich Informatik, TU-Berlin, 1994.
- [15] W.G. Kropatsch, M. Eder, P. Kammerer, "Finding strokes of the brush in portrait miniatures", in: N. Pavesic, H. Niemann, S. Kovacic, F. Mihelic (Eds.), "*Speech and Image Understanding, Proc. of the 3rd Slovenian-German Workshop and 2nd SDRV Workshop, Ljubljana*", pp.257-265, 1995.
- [16] X. Jia, M.S. Nixon, "Extending the feature vector for automatic face recognition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.17, No.12, pp.1167-1176, 1995.
- [17] H. Levkowitz, G. T. Herman, "Glls: A generalized lightness, hue, and saturation color model", *Computer Vision, Graphics, and Image Processing*, Vol.55, No.4, pp.271-285, 1993.
- [18] X. Li, N. Roeder, "Face contour extraction from front-view images", *Pat. Rec.*, Vol.28, No.8, pp.1167-1179, 1995.
- [19] T. Lindeberg, "*Discrete scale space theory and the scale-space primal sketch*", PhD Thesis, Computational Vision and Perception Laboratory (CVAP), Stockholm, Sweden, 1991.
- [20] D. Marr, E. Hildreth, "Theory of edge detection", *Proc. R. Soc. Lond.*, pp.187-217, 1980.
- [21] G.M. Murch, S.M. Taylor, "*Color in computer graphics: manipulation and matching colors*", Springer Verlag, Berlin, Heidelberg, New York, 1989.
- [22] V.S. Nalwa, "*A guided tour of computer vision*", Addison-Wesley Publishing Company, 1993.
- [23] T. Nemeč, "*Gesichtserkennungssystem zur Raumüberwachung*", Master's thesis, Vienna Univ. of Technology, 1995.
- [24] A. Rosenfeld, A.C. Kak, "*Digital picture processing*", Vol.2, Academic Press, Inc., 1982.
- [25] A. Rosenfeld, "Axial representations of shape", *Computer, Vision, Graphics and Image Processing*, Vol.33, pp.156-173, 1986.
- [26] M.W. Schwarz, W.B. Cowan, J. C. Beatty, "An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models", *ACM Transactions on Graphics*, Vol.6, No.2, pp.123-158, 1987.
- [27] I. Tastl, R. Sablatnig, W.G. Kropatsch, "Model-based classification of painted portraits", in: A. Pinz, (Ed.), "*Pattern Recognition 1996, Proc. of the 20th ÖAGM- Workshop*", Schriftenreihe der OCG, Vol.90, pp.237-249, Oldenburg, Wien, München, 1996.

Chapter 6

Model-based Registration of Front- and Backviews (“Registration”)

Robert Sablatnig and Martin Kampel. Model-based Registration of Front- and Backviews. *Computer Vision and Image Understanding*, 87(1):90–103, 2002.

Model-based Registration of Front- and Backviews of Rotationally Symmetric Objects^{*}

Robert Sablatnig and Martin Kampel

*Pattern Recognition and Image Processing Group,
Institute of Computer Aided Automation,
Vienna University of Technology,
Favoritenstraße 3/183/2, A-1040 Vienna, Austria*

Abstract

This paper shows an algorithm that pre-aligns the front- and the backviews of rotationally symmetric objects for the registration of the two 3d-surfaces without using corresponding points. The geometric alignment of the two three-dimensional surfaces is then performed by using a modified ICP (Iterative Closest Point) algorithm, which needs an initial estimate of the relative pose. The method proposed uses the axis of rotation of fragments to bring two range images into alignment. We are developing a classification system for archaeological fragments based on their profile, which is the cross-section of the fragment in the direction of the rotational axis of symmetry. Hence, the correct registration of the front- and backview are important. We demonstrate the method and give results on synthetic and real data.

Key words: Range Image Registration, Pose Estimation, ICP, Ceramics, Archaeology.

1 Introduction

The typical 3d-scanner output are range images from objects from one direction at a time. These range images have then to be registered to one another in order to reconstruct the complete object in 3d. Normally, this registration is

^{*} 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.

purely based on geometry since color is not always provided by the range scanner and may also differ viewed from different directions. The most prominent and dominant method for aligning three dimensional surfaces is the ICP (Iterative Closest Point) algorithm introduced by Chen and Medioni [1] and Besl and McKay [2]. Since then, many variants have been introduced on the basis of ICP, some of these variants (like Rusinkiewicz and Levoy [3]) expand the abbreviation also to Iterative Corresponding Point claiming that this would better suit the algorithm.

ICP starts with two surface meshes and an initial guess for their relative rigid-body transform and iteratively refines the transform by generating pairs of corresponding points and minimizing an error metric [3]. Generating the initial alignment is a critical issue within this context and is application dependent. It can be based on the scanner position, surface features, exhaustive search for corresponding points [4] or interactive user input (which is by far the most prominent method [5]). All the initial alignment and ICP methods rely on - at least partially - overlapping range images. When handling with very specific objects like fragments of ancient ceramics, 3d-range images do not provide overlapping regions in the range images. In such cases model based approaches to the problem have to be made, which narrow the range of applicable objects to be registered.

In this paper we present a method to solve the 3d-registration problem of front- and backview for rotationally symmetric objects. The 3d-shape of ceramic fragments is computed using these two views in order to virtually reconstruct the complete vessel (where the fragment is a part of). Pottery is assumed to be rotationally symmetric with only one axis of rotation since it was made on a rotation plate. To perform the registration of the two surfaces of one fragment, we use a-priori information about fragments belonging to one vessel: both surfaces have the same axis of rotation since they belong to the same object. With respect to this property the axis of rotation is calculated using a Hough inspired method [6]. In this paper we concentrate on the registration of the front- and backview of one fragment, which is significantly different from registering the surfaces of different fragments of one object in order to reconstruct the object out of its pieces.

Figure 1 gives an overview of the 3d-surface reconstruction from two object views and basically shows the structure of this paper. In Section 2 a short introduction to our application is given to motivate the registration process. The first step consists of sensing the front- and backside of the rotationally symmetric fragment using a calibrated 3d-acquisition system. We use a Shape from Structured Light method based on active triangulation. The images obtained are computed and the result is a 2D array of depth values (see [7] for details on the acquisition system). The resulting range images are used to estimate the axes of rotation, shown in Section 3. Section 4 presents the proposed

registration method for the surface reconstruction and results are presented in Section 5. We conclude the paper with a discussion of the results and give an outlook on future work.

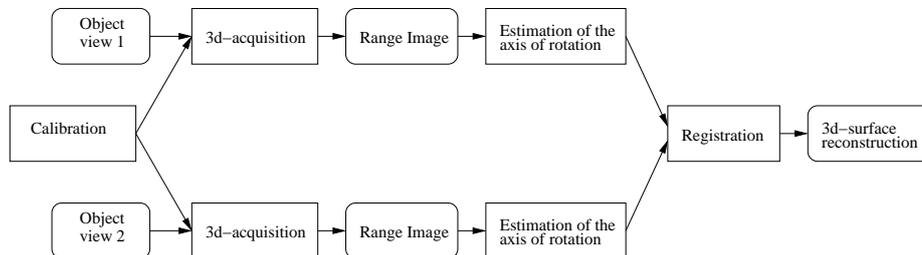


Fig. 1. Overview on 3d-reconstruction from two object views.

2 Motivation - Archaeological Background

In this paper we address a very specific problem on registering front- and backsides of rotationally symmetric objects, namely fragments of ancient ceramics. 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. At excavations a large number of ceramic fragments, called sherds are found. These fragments are photographed, measured, drawn (called *documentation*) and classified. Figure 2 shows such an example (taken from [8]) of a drawing of a pot. On the right hand side the profile is shown (black part of the figure, four parallel lines on the top show that there are rills on the top of the pot), the rest of the figure shows the decoration (rills) on the surface of the pot and its rotational axis. Therefore, the important properties of such a drawing are outer and inner profile shape, diameter and surface characteristics

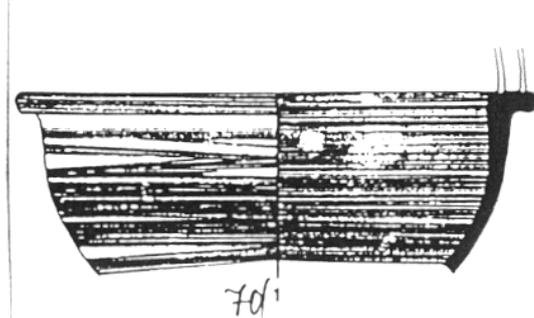


Fig. 2. Drawing of a complete pot (from [8])

Traditional archaeological classification is based on the so-called profile of the object, which is the cross-section of the fragment in the direction of the rotational axis of symmetry. This two-dimensional plot holds all the information needed to perform archaeological research. The correct profile and the correct axis of rotation are thus essential to reconstruct and classify archaeological ceramics.

Because the conventional method for documentation is unsatisfactory [9], we developed an automated 3d-object acquisition system with respect to archaeological requirements [10].

3 Estimation of Axes of Rotation

The basis for this axis estimation is a dense range image provided by the range sensor. The estimation approach exploits the fact, that surface normals of rotationally symmetric objects intersect their axis of rotation. If we have an object of revolution (like an archaeological vessel made on a rotation plate) we can suppose that all intersections of the surface normals n_i are positioned along the axis of symmetry R , which is schematically shown in Figure 3.

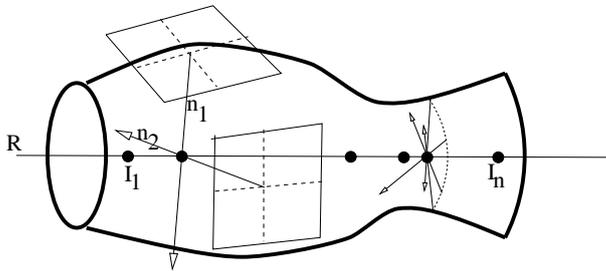


Fig. 3. Axis determination using Hough-Space.

This assumption holds [11] for a complete object or even for its fragment. We try to extract a surface of revolution where the curvature of the surface is relatively small. A few approaches to extract volumetric shape descriptors of solids of revolution out of dense range images are reported in Yokoya and Levine [12] and a Hough-based approach to the problem is presented. The Hough transform is a robust and efficient tool for feature extraction [13,14]. It is based on a voting principle: each point or element will vote for the set of features to which it could belong. This voting principle makes the Hough transform very robust toward noise or outliers [15]. Yokoya and Levine [12] use the center of the principal curvature from first and second partial derivatives of the surface, which construct the so-called focal surface (see [16]). Since our surfaces have a relatively small curvature we adopted this method by using a robust way for the determination of the surface normals, based on planar patches.

We consider a planar patch of size $s \times s$. The patch is fitted according to the following equation:

$$ax + by + cz + d = 0 . \quad (1)$$

This defines a planar patch with normal $\vec{n} = (a, b, c)$. The fitting algorithm used is the Total Least Squares (TLS). Let $\mathbf{X} = \{ \mathbf{X}_1, \dots, \mathbf{X}_N \}$, where $\mathbf{X}_i =$

(x_i, y_i, z_i) . The TLS minimizes the following expression

$$E_n = \sum_{i=1}^N r_i^2, \text{ where } r_i = \frac{|ax_i + by_i + cz_i + d|}{\sqrt{a^2 + b^2 + c^2}}. \quad (2)$$

It was shown that this approach is equivalent to the MCA or Minor Component Analysis [17]: $\min_{\vec{\mathbf{n}}} E_n \equiv \mathbf{A} \cdot \vec{\mathbf{n}} = \lambda_{min} \vec{\mathbf{n}}$, where \mathbf{A} is the covariance matrix of set \mathbf{X} and λ_{min} is the smallest eigenvalue. The constant d is determined by

$$d = -\vec{\mathbf{n}} \cdot \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i. \quad (3)$$

The main goal is to minimize the distances between the points of the surface and the planar patch. An iteratively reweighted algorithm is used to compute the optimal value of the normal and discard outliers. The objective of the algorithm is to achieve:

$$\Delta = \min_{\vec{\mathbf{n}}} \sum_{i=1}^{s^2} [ax_i + by_i + cz_i + d]^2. \quad (4)$$

In order to minimize the function, we use an iterative scheme. The points are weighted according to their residual [18], a point M_i at iteration k and with residual r_i has a weight ω_k defined by:

$$\omega_k = \begin{cases} 1 & \text{if } r_i \leq Sa \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where S is the Median Absolute Deviation and a is a tuning constant. The algorithm can be outlined with the initial state $k = 0$, $\omega_0 = [1 \dots 1]$ as follows:

- (1) compute the surface normal $\vec{\mathbf{n}}_k$ using weights w_k .
- (2) compute the residuals r_i with the estimated parameters.
- (3) compute ω_{k+1} based on r_i .
- (4) iterate steps 1., 2., and 3 until convergence.

The algorithm uses all data points in the initial step, since it is assumed that there are only a few outliers in the data and the number of data-points within the patch is relatively low (around 20). If these conditions do not hold this produces a wrong surface normal that is eliminated by the subsequent axis determination.

Once the surface normals for all points are computed the rotation axis R can be estimated. 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 R 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 the points in the accumulator with a high counter value are defined as maxima. These maxima form the axis of rotation.

The accumulator maxima are taken as candidate points for the estimation of the rotation axis. For the set of points $\text{acc}(x, y, z)$ a PCA is used to find the optimal axis going through these set of points \mathcal{M} with

$$\mathcal{M} = \{P_i(x, y, z) | g_i = \text{acc}(x, y, z) > T_{acc}\}, \quad (6)$$

with

$$T_{acc} = s * \max(\text{acc}), \quad (7)$$

where T_{acc} defines a threshold accepting points in the accumulator. In order to estimate the axis the vector $\vec{\mathbf{v}}$ is determined by $\mathbf{A}\vec{\mathbf{v}} = \lambda_{max}\vec{\mathbf{v}}$, where \mathbf{A} is the covariance matrix of set \mathcal{M} and λ_{max} is the largest eigenvalue of \mathbf{A} .

A point \mathbf{G} on the axis is determined by using all points in \mathcal{M} with

$$G_{\mathcal{M}} = \frac{\sum_{i \in \mathcal{M}} \tilde{g}_i \mathbf{P}_i}{\sum_{i \in \mathcal{M}} \tilde{g}_i}, \text{ where } \tilde{g}_i = \begin{cases} g_i & d_i \leq T_d \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

and the threshold $T_d = aS$. With the point \mathbf{G} and the vector $\vec{\mathbf{v}}$ the rotation axis is defined, thus a robust version of the complete algorithm can be outlined. The sets \mathcal{I}_k and \mathcal{O}_k are respectively the inlier and outlier sets and the method starts with the initial condition $k=0$ and $\mathcal{I}_k = \mathcal{M}$ and $\mathcal{O}_k = \{\}$.

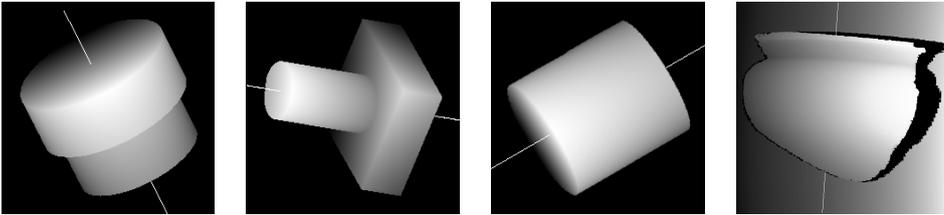
- (1) compute surface normals $\vec{\mathbf{n}}$ for all points of the object.
- (2) cluster lines \mathcal{L}_i in $\text{acc}(x, y, z)$.
- (3) compute $\vec{\mathbf{v}}_k$ and $\mathbf{G}_{\mathcal{I}_k}$ using set \mathcal{I}_k .
- (4) determine the distances d_i of set \mathcal{M} to the axis defined by $\mathbf{G}_{\mathcal{I}_k}$ and $\vec{\mathbf{v}}_k$.
- (5) update \mathcal{O}_{k+1} , $\mathcal{I}_{k+1} = \mathcal{M} - \mathcal{O}_{k+1}$.
- (6) iterate steps 3., 4. and 5. until convergence or a maximum number of iteration is reached.

Using this technique outliers introduced by noisy range data, based on a bad calibration or discretization errors, can be avoided, since in the Hough-Space wrong data points are in the minority and do not build a maximum. To evaluate this, synthetic range images are used where the axis of rotation is known and the images are disturbed by non symmetric object parts. Since there are some threshold parameters, the affect of changing these parameter was also taken into consideration.

The size of s of the planar patches depends on the geometry of the object and the accuracy of the range sensor, the more noise is expected the larger s should be to eliminate the outliers. However, since we are considering curved

surfaces s is also influenced by the minimal curvature we want to estimate in relation to the sensor resolution. In our tests we used different sizes it showed to work best with $s = 5$, which ensures to have at least 20 range points within the patch. The tuning constant was set to $a = 1$ (and therefore not used) and the threshold $T_{acc} = 0.7$ in all of the experiments, both parameters were determined empirically based on our test data. The parameter space of the hough space was set to $400 \times 400 \times 800$, which corresponds to the range image size of 400×400 and the maximal possible diameter of the fragment. The computation time depends on the number of range points (size of the object) and varies between 5min and 15 min on a Pentium 233MMX, 256 MB RAM using non-optimized code.

In order to determine the error of the axis computation, the Mean Square Error (MSE) between the original and the computed axis is determined. The MSE is computed for all points of the axis inside the test object. The MSE of all distances between the estimated and correct axis define the error. In Figure 4 test objects are visualized, where the estimated axis of rotation is computed into the range image. For each test image the MSE is given according to the position of the original axis. Compared to simple least square solutions only one wrong surface normal, which is sufficiently far away from the bulk of data can ruin the estimation completely. It can be shown that using our method the axis can be determined even if there are large regions in the range image, which are not rotationally symmetric as can be seen in Figure 4b. In our tests we found out that in the average 15% of the surface normals were wrong, which did not (or only slightly) influence the accuracy of the axis determination since these normals were eliminated by the Hough method.



(a) $MSE=0.16$ (b) $MSE=0.20$ (c) $MSE=0.11$ (d)

Fig. 4. Axis determination for synthetic and real range data.

Problems that arise with real data are symmetry constraints, i.e. if the surface of the fragment is too flat or too small, the computation of the rotational axis is ambiguous (worst case: sphere) resulting in sparse clusters in the Hough-space, which indicate that the rotational axis is not determinable. Therefore, before we start our registration algorithm we first compute the Gaussian curvature for 20% of randomly selected surface points of one surface (back side, since there are usually no decorations on). We determine whether the curvature is large enough to estimate the rotational axis, i.e we look at curves determined by the intersection between the surface and planes perpendicular to the tangent plane at each surface point. All these curves have a single and well defined

curvature at the point. The maximum and minimum of the normal curvature κ_1 and κ_2 at a given point on a surface are called the principal curvatures. The principal curvatures measure the maximum and minimum bending of a regular surface at each point. The Gaussian curvature K and mean curvature H are related to κ_1 and κ_2 by: $K = \kappa_1 * \kappa_2$ and $H = 1/2(\kappa_1 + \kappa_2)$. We introduce two thresholds - a lower and an upper - to evaluate the results of the axis determination. If the average mean curvature of all selected surface points is below a lower threshold (0.2) the process is not started and the fragment is marked uncomputable. If it is above the lower threshold and below the upper threshold (0.8) it is marked computable with a confidence weight, and if it is above the threshold it is again very unlikely to get good results since the surface is sphere-like (which is almost never the case in our test data). For both non-computing cases, other registration strategies have to be used, for flat fragments the surfaces are registered assuming that they are completely flat and should match to one another, for sphere-like fragments no registration is performed since these fragments do not hold any relevant archaeological information (they are not orientable manually as well). Figure 4d shows the result for a frontview of a fragment with the estimated rotational axis (black regions in Figure 4d indicate points where no range information is available due to occlusion).

4 Range Image Registration

The task of building full 3d-models of general objects is difficult, since there is no a-priori knowledge about the shape of the object. One simple method is to use a calibrated turntable upon which the camera is fixed, as described in [19]. Even though the turntable method described above is good at creating 3d-models, there is still the question of getting the bottom of the object sitting on the turntable. So the bottom and the top of the object needs to be scanned in and then registered. In the case of thin ceramic fragments, the rotation table method does not solve the registration problem since one view of the fragment is always invisible (one solution would be to "glue" the fragment onto the plate in an upright position, however, this method is impractical and unthinkable for archaeologists [20]).

Fragments of vessels are thin objects, therefore 3d-data of the edges of fragments are not accurate and this data can not be acquired without placing and fixing the fragment manually. Ideally, the fragment is placed in the measurement area, a range image is computed, the fragment is turned and again a range image is computed. To perform the registration of the two surfaces, 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. Furthermore, 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 in the average. Note that this is only the case if the profile is taken perpendicular to the axis since this is induced by the manufacturing process. This condition holds for almost all fragments with the exception of relief decorated fragments. Still the average distance perpendicular to the rotational axis is constant on most of the parts of the fragments since there are more un-decorated than decorated parts. Therefore, this assumption is used to perform the "fine registration".

The most commonly used algorithms for registering is the ICP algorithm [5]. 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 algorithm starts with the selection of some point sets in one or both surfaces (which generally are triangulated surfaces), matches these point sets to one another, which gives a set of corresponding pairs, and weights the corresponding pairs. A rejection rule for pairs is applied to all pairs to determine outliers. To measure the fit, an error metric is used, which is minimized iteratively.

There are many different variants of the ICP Algorithm (see [3] for a review) all based on local point correspondences. Therefore, it is very important to have a good rough alignment of the surfaces to be registered. Algorithms that do not use single pair of surface registration (no pre-alignment) are also called *global registration* algorithms (see [21], [22], or [23] for details).

Since we have a-priori knowledge about our surfaces and the rotational axis estimation we decided to use a computationally relatively inexpensive model-based approach. No point to point correspondences are required to determine the interframe transformation needed to express the points from each view in a common reference coordinate system [24]. We register the range images by calculating the axis of rotation of each view (Figure 5a and Figure 5b) and by bringing the resulting axes into alignment (Figure 5c). 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 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. Note that after the rough alignment (vertical and horizontal) due to inaccurate estimation of the rotational axis the two surfaces may intersect (Figure 5d).

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 [21]. 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 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.

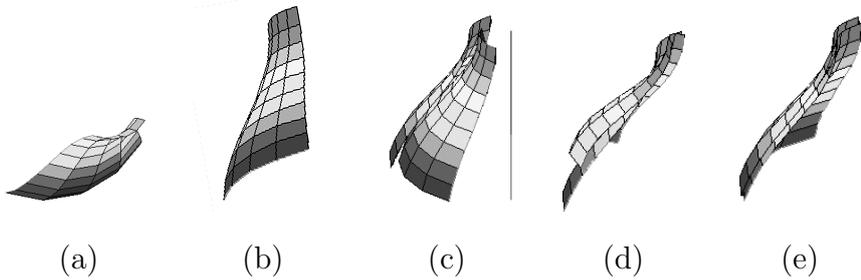


Fig. 5. Registration steps using synthetic data.

The first estimation of all d_n is given by the range sensor. Since both range images are computed in the same environment and the fragments are placed on a plane, the range sensor computes the normal distance of the inner surface to the object plane on the lowest point of the surface - which is an estimation of the thickness of the fragment in a stable position.

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. To sum up, the registration algorithm can be outlined as follows:

- (1) compute the axis of rotation for each view.
- (2) compute the vertical alignment by top points of each view.
- (3) compute the horizontal alignment until the maximum number of points in common projection is reached.
- (4) set all d_n to the same initial value given by the range sensor.
- (5) compute all actual distances d_n .
- (6) compute all errors δ_i .
- (7) iterate steps 5 and 6 until all δ_i are positive
- (8) update all d_n to new mean distance in direction of rotational axis.
- (9) iterate steps 5, 6, and 8 until all δ_i are minimal or MSE δ increases.

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

5 Results

We tested our method on synthetic range images of a synthetic fragment (thickness 2.9mm) with approx. 7000 surface points each where we had a registration error $\delta=0.02\text{mm}$ (see Table 1: synth 2). In comparison to the previous results the registration error is smaller since there are more surface points and therefore the computation of the rotation axis is much better.

To find out if the method is working on real data we used a totally symmetric small flowerpot with known dimensions and took a fragment which covered approximately 25% of the original surface. The results are given in Table 1 for real 4. The distribution of the registration error δ for the flowerpot is shown in Figure 6a, where the registration error δ (in mm) is shown as line in the z -direction of the coordinate system, x - and y -coordinates denote the fragment's x - and y -dimensions. The registration error increases towards the top of the pot, because of the irregularity of the distance between the surfaces at that region since the flowerpot has an edge (upper border) where inner and outer surface are not parallel due to misalignment of the surfaces.

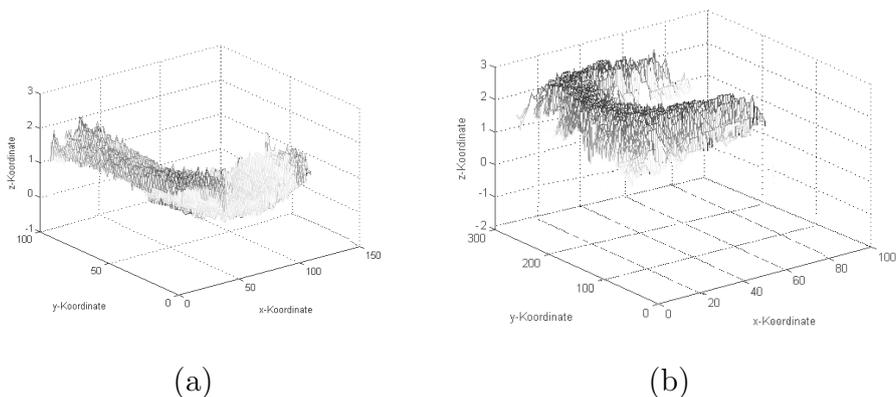


Fig. 6. Distribution of δ : registered flowerpot (a) archaeological fragment (b).

Figure 7c and d show the frontview, backview and the axis of rotation of a real archaeological fragment. Results of the registration tests with this fragment (real 16) are shown in Table 1. Figure 8 shows the registration of intersecting surfaces for real data in detail: Figure 8a and Figure 8b show intersecting surfaces due to wrong rotational axis estimation, Figure 8c shows the same surfaces after the ICP-based registration procedure.

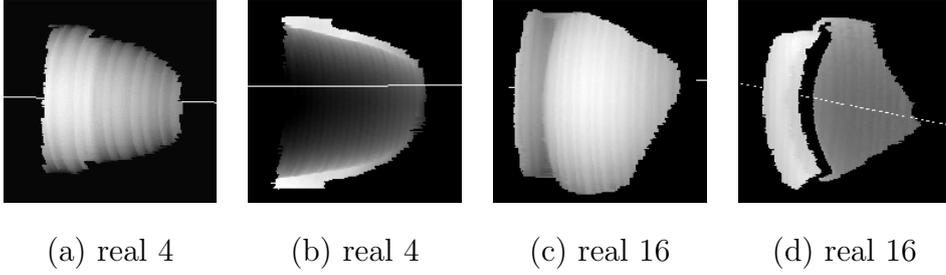


Fig. 7. Front- and backview and their axis of rotation of a flowerpot real 4 (a,b) and an archaeological fragment real 16 (c,d).

Figure 6b shows the distribution of δ of a registered archaeological fragment. Marginal peaks are caused by shadow regions of the backview (see (Figure 7d) at the border of the fragment, where either no range data is processed or the range information is unreliable. The increase of the registration error δ reflects the uneven surface of the fragment.

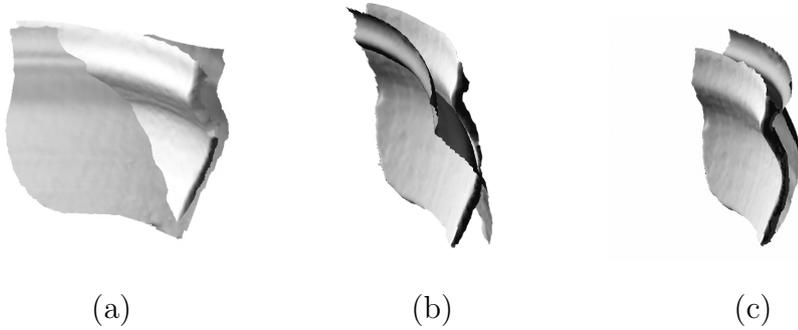


Fig. 8. Registration steps using real data.

Table 1 gives an overview of the presented results. It shows the number of points of the back- and frontview, the mean curvature H , the mean thickness of the fragment d , the estimated registration error δ , and the relative error δ/d . The increase of δ between the synthetic and real data tests is caused by the error in the determination of the rotational axis.

The evaluation of the results 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). Figure 9 shows the correlation (0.83) between the difference of number of points between the two views (solid line, left vertical axis) and the relative registration error (dotted line, right vertical axis). The horizontal axis denotes the samples (data type) as given in Table 1.

The next parameter that influences the results is the mean mean curvature H . Since the registration algorithm uses the axis of rotation for rough alignment, surfaces that are flat cannot be registered since the rough alignment does not work. In our tests no results were obtained if $H < 0.2$ due to completely wrong axis estimations.

Data type	# of points (back view)	# of points (front view)	H	error δ [mm]	mean d [mm]	rel. error δ/d
synth. 1	50	50	0.53	0.05	2.9	0.0172
synth. 2	6674	6674	0.47	0.02	2.9	0.0069
real 1	52324	52350	0.72	0.31	5.5	0.0564
real 2	46800	46602	0.61	0.49	7.9	0.0620
real 3	28745	29210	0.68	0.30	4.8	0.0625
real 4	10191	9619	0.24	0.44	5.6	0.0786
real 5	21990	22564	0.42	0.60	6.1	0.0984
real 6	24249	25021	0.41	0.29	3.2	0.0906
real 7	40870	41750	0.46	0.69	6.8	0.1015
real 8	48320	49433	0.36	0.53	4.8	0.1104
real 9	18307	17048	0.59	0.65	5.2	0.1250
real 10	60436	62173	0.39	0.74	6.0	0.1233
real 11	28340	30247	0.49	0.49	3.7	0.1324
real 12	49133	47051	0.64	1.35	8.7	0.1552
real 13	136342	134220	0.44	1.37	8.9	0.1539
real 14	142856	147174	0.34	1.64	9.5	0.1726
real 15	44587	39877	0.47	1.11	6.4	0.1734
real 16	31298	37176	0.27	1.19	6.9	0.1725
real 17	27282	33571	0.57	1.06	5.9	0.1797
real 18	71199	81735	0.53	1.98	8.7	0.2276
real 19	32124	21214	0.58	0.99	4.2	0.2357
real 20	120761	134319	0.32	2,59	10.3	0.2515

Table 1
Results of the registration process.

One source of error lies inherently in the algorithm since in the "fine registration" process one surface (backview) is fixed and the other surface is registered in relation to the fixed surface. Therefore the error made in the estimation of the axis of rotation of the backview is not corrected and influences the registration result.

From the results one can see that the maximum error was approximately 25%, which is acceptable in this specific application only since the main goal is to

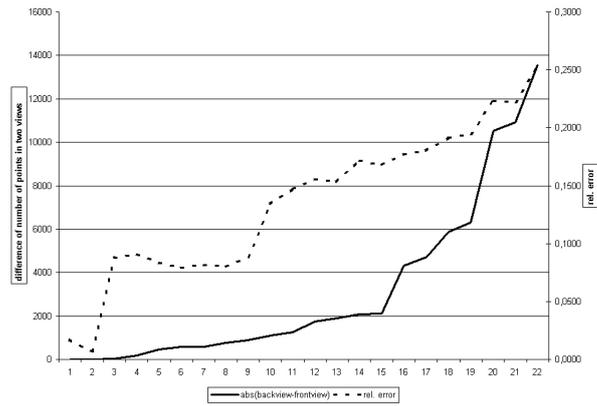


Fig. 9. Correlation between difference of number of points in the two views and the registration error.

compute the profile and the outer profile is the most important attribute. Error measures are also given to the user so that it is possible to manually correct the result if necessary. However, the average error of 18% was accepted by the archaeologists since manual drawings and measurement have more errors [9] (in fragment thickness terms this is 1mm error in the profile orientation if the fragment is 4mm thick, which is low in manual drawing). The rendered 3d-models are also used for visualisation as can be seen in Figure 10. Figure 10a shows the fragment real 6 from 1, Figure 10b and c show 2 more fragments as examples.

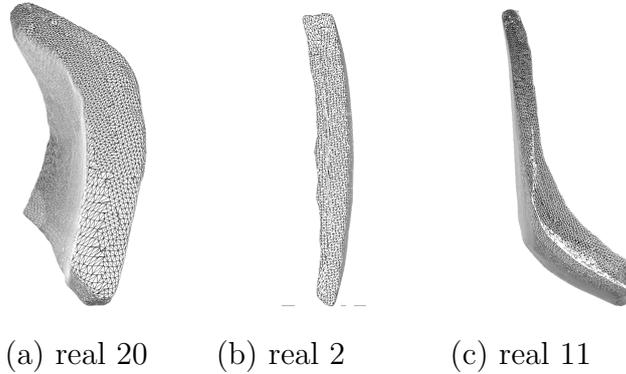


Fig. 10. Visual results for registered fragments.

6 Conclusion and Outlook

We have proposed a pre-alignment algorithm for registering the front- and backview of rotationally symmetric objects from range data. The work was performed in the framework of the documentation of ceramic fragments. For this kind of objects, the surfaces to be registered have to be pre-aligned carefully because otherwise pair-wise registration techniques fail, since there are no corresponding points in the range images. We demonstrated a model based

technique that computes and uses the axis of rotation of fragments belonging to the same vessel to bring two views of a scene into alignment. We used a robust technique to determine both surface normals and the rotational axis.

The method has been tested on synthetic and real data with reasonably good results. It is part of continuing research efforts to improve the results from various range images since the technique depends on the correct determination of the rotational axis of one surface. Furthermore we want to conduct intensive tests with real archaeological fragments within the 3D-MURALE project [25] that are selected, provided, and evaluated by archaeologists.

References

- [1] Y. Chen, G. Medioni, Object Modelling by Registration of Multiple Range Images, *Image and Vision Computing* 10 (3) (1992) 145–155.
- [2] P. Besl, N. McKay, A Method for Registration of 3-D Shapes, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14 (2).
- [3] S. Rusinkiewicz, L. M., Efficient Variants of the ICP Algorithm, in: *Proc. of the 3rd IEEE Intl. Conference on 3-D Digital Imaging and Modeling, Quebec, 2001*, pp. 145–152.
- [4] C. Chen, Y. Hung, J. Cheng, RANSAC-Based DARCES: A New Approach to Fast Automatic Registration of Partially Overlapping Range Images, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 21 (11) (1999) 1229–1234.
- [5] M. Greenspan, G. G., A Nearest Neighbor Method for Efficient ICP, in: *Proc. of the 3rd IEEE Intl. Conference on 3-D Digital Imaging and Modeling, Quebec, 2001*, pp. 161–168.
- [6] S. B. Yacoub, C. Menard, Robust Axis Determination for Rotational Symmetric Objects out of Range data, in: W. Burger, M. Burge (Eds.), *21 th Workshop of the Oeagm, Hallstatt, Austria, 1997*, pp. 197–202.
- [7] M. Kampel, R. Sablatnig, On 3d Modelling of Archaeological Sherds, in: N. Sarris, M. Strinzis (Eds.), *Proc. of Intl. Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, Santorini, 1999*, pp. 95–98.
- [8] P. Kenrick, Rim-forms of some Relief-decorated Vessels in Italian Terra Sigillata, *Conspectus formarum terrae sigillatae italico modo confectae*, Bonn, 1990.
- [9] C. Orton, P. Tyers, A. Vince, *Pottery in Archaeology* (1993).
- [10] M. Kampel, R. Sablatnig, Automated 3D recording of archaeological pottery, in: D. Bearman and F. Garzotto (Ed.). *Proc. of the Int. Conf on Cultural Heritage and Technologies in the Third Millennium Milan, Italy, 2001*, pp. 169–182.

- [11] R. Halíř, Estimation of Rotation of Fragments of Archaeological Pottery, in: W. Burger, M. Burge (Eds.), 21 th Workshop of the Oeagm, Hallstatt, Austria, 1997, pp. 175–184.
- [12] N. Yokoya, M. Levine, Volumetric Shapes of Solids of Revolution from a Single-View Range Image, *Computer Vision, Graphics, and Image Processing* 59 (1) (1994) 43–52.
- [13] R. Yip, W. Lam, P. Tam, D. Leung, A Hough Transform Technique for the Detection of Rotational Symmetry, *Pattern Recognition Letters* 15(9) (1994) 919–928.
- [14] M. Hoffelder, K. Sauer, J. Rigby, A Hough Transform Technique for Detection of Rotationally Invariant Surface Features, in: *Proc. of 1st. Intl. Conference on Image Processing*, Austin, Texas, 1994, pp. 944–948.
- [15] D. Ballard, C. Brown, *Computer Vision*, Prentice Hall, 1982.
- [16] N. Yokoya, M. Levine, Volumetric Descriptions of Solids of Revolution in a Range Image, in: *Proc. of 10th. Intl. Conference on Pattern Recognition*, Atlantic City, NJ, 1990, pp. 303–308.
- [17] L. Xu, E. Oja, C. Suen, Modified Hebian Learning for Curve and Surface Fitting, *Neural Networks* 5 (3) (1992) 441–457.
- [18] P. Huber, *Robust Statistics*, Wiley, New York, 1981.
- [19] R. Szeliski, Image Mosaicing for Tele-Reality Applications, in: *Proc. of 2nd. Workshop on Applications of Computer Vision*, Sarasota, 1994, pp. 44–53.
- [20] S. Tosovic, R. Sablatnig, 3d Modeling of Archaeological Vessels using Shape from Silhouette, in: *Proc. of the 3rd IEEE Intl. Conference on 3-D Digital Imaging and Modeling*, Quebec, 2001, pp. 51–58.
- [21] K. Pulli, Multiview Registration for Large Data Sets, in: *Proc. of the 2nd IEEE Intl. Conference on 3-D Digital Imaging and Modeling*, Ottawa, 1999, pp. 160–168.
- [22] C. Dorai, G. Wang, A. Jain, C. Mercer, Registration and Integration of Multiple Object Views for 3D Model Construction, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 20 (1) (1998) 83–89.
- [23] J. Williams, M. Bennamoun, A Multiple View 3D Registration Algorithm with Statistical Error Modeling, *IEICE Trans. Inf. and Syst.* E83-D (8) (2000) 1662–1670.
- [24] B. Vemuri, J. Aggarwal, 3D Model Construction from Multiple Views using Range and Intensity Data, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1986, pp. 435–437.
- [25] M. Kampel, R. Sablatnig, On using 3d multimedia tools to measure, reconstruct and visualize an archaeological site, *Forum Archaeologiae, Zeitschrift fuer klassische Archaeologie* 19 (IV).

Chapter 7

Combining Shape from Silhouette and Shape from Structured Light for Volume Estimation of Archaeological Vessels (“Combination”)

Robert Sablatnig and Srdan Tosovic and Martin Kampel. Combining Shape from Silhouette and Shape from Structured Light for Volume Estimation of Archaeological Vessels. In R. Kasturi, D. Laurendeau, and C. Suen, editors, *Proc. of 16th International Conference on Pattern Recognition, Quebec City*, volume 1, pages 364–367. IEEE Computer Society, 2002.

Combining Shape from Silhouette and Shape from Structured Light for Volume Estimation of Archaeological Vessels *

Robert Sablatnig, Srdan Tosovic, and Martin Kampel
Vienna University of Technology,
Institute of Computer Aided Automation,
Pattern Recognition and Image Processing Group
Favoritenstr. 9, 183-2, A-1040 Vienna, Austria
{sab,kampel}@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 two 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* (SfL) method presented in this paper was performed within the *Computer Aided Classification of Ceramics* [6] project, which aims to provide an objective and automated method for classification and reconstruction of archaeological pottery. The volume of the vessel is of interest to archaeologists, since the volume estimation allows a more precise classification.

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 [2, 12]. 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* [7]. The method can be applied on objects of arbitrary shapes, including objects with certain concavities, as long as the concavities are visible from at least one input view [14]. This condition is hard to hold since most of the archaeological vessels do have concavities that have to be modeled. Therefore, a second, active shape determination method has to be used to discover all concavities. The second acquisition method used is SfL, based on active triangulation [3, 4].

There have been many works on construction of 3d models of objects from multiple views. Baker [1] used silhouettes of an object rotating on a turntable to construct a wire-frame model of the object. Martin and Aggarwal [10] constructed volume segment models from orthographic projection of silhouettes. Potmesil [12] created octree models using arbitrary views and perspective projection. In contrast to this, Szeliski [13] first created a low resolution octree model quickly and then refined this model iteratively, by intersecting each new silhouette with the already existing model. The work of Szeliski [13] and Niem [11] was used as a basis for the SfS approach presented in this paper. For the active triangulation method we use an approach by Liska developed for a next view planing strategy using structured light [9].

2. Acquisition System

The acquisition system consists of the following devices:

- a turntable (Figure 1a) with a diameter of 50 cm, and a positional accuracy of 0.05° .
- two monochrome CCD-cameras with a focal length of 16 mm and a resolution of 768x576 pixels. One camera (*Camera-1* in Figure 1) is used for acquiring the images of the object's silhouettes and the other (*Camera-2* in Figure 1) for the acquisition of the images of the laser light projected onto the object.

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

- a red laser (Figure 1d) used to project a light plane onto the object. The laser is equipped with a prism in order to span a plane out of the laser beam.
- a lamp (Figure 1e) used to back-light [5] the scene for the acquisition of the silhouette of the object.

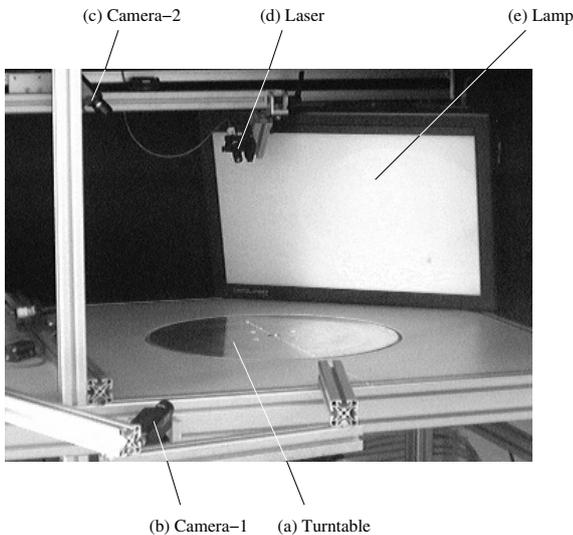


Figure 1. Acquisition system.

Both cameras are placed about 50 cm away from the rotational axis of the turntable. Ideally the optical axis of the camera for acquiring object's silhouettes 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 views the turntable from an angle of about 45°. The laser is directed such that the light plane projected contains the rotational axis of the turntable. Prior to any acquisition, the system is calibrated in order to determine the inner and outer orientation of the cameras and the rotational axis of the turntable (for details see [14]).

3 Combination of Algorithms

An input image for SfS defines a conic volume in space which contains the object to be modeled (Figure 2a). Another input image taken from a different view defines another conic volume containing the object (Figure 2b). Intersection of the two conic volumes narrows down the space the object can possibly occupy (Figure 2c). With an increasing number of views the intersection of all conic volumes approximates the actual volume occupied by the object better and better, converging to the 3d visual hull of the object [8]. Therefore by its nature SfS defines a *volumetric* model of an object.

An input image for SfL using laser light defines solely the points on the surface of the object which intersect the laser plane (Figure 3a). Multiple views provide a cloud of

points belonging to the object surface (Figure 3b), which is a *surface* model of the object.

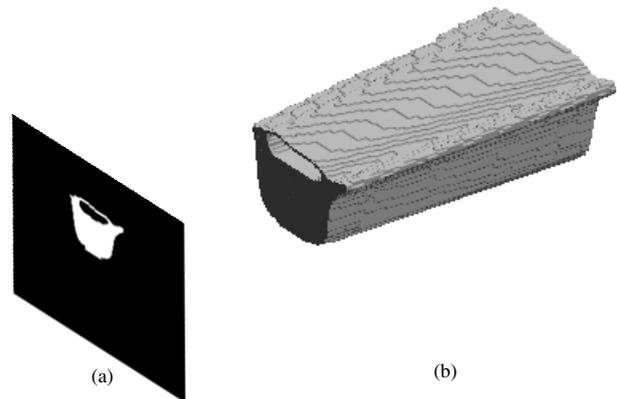


Figure 2. Two conic volumes and their intersection.

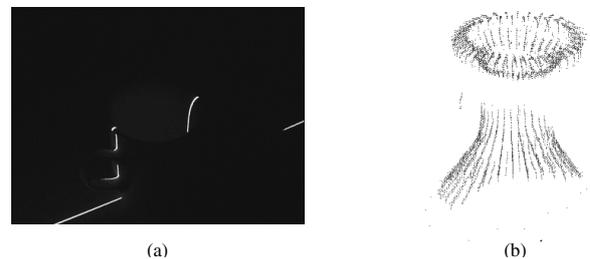


Figure 3. Laser projection and cloud of points.

The main problem to be addressed in an attempt to combine these two methods is how to adapt the two representations to one another, i.e. how to build a common 3d model representation. One possibility would be to build a separate SfL surface model and a SfS volumetric model followed by converting one model to the other and intersecting them. But if we want to estimate the volume of an object using our model, any intermediate surface models should be avoided because of the problems of conversion to a volumetric model. Therefore, our approach proposes building a single volumetric model from the ground up, using both underlying methods (see Figure 4):

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 SfL.

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. In addition to the images of the object (Figure 4a, 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.

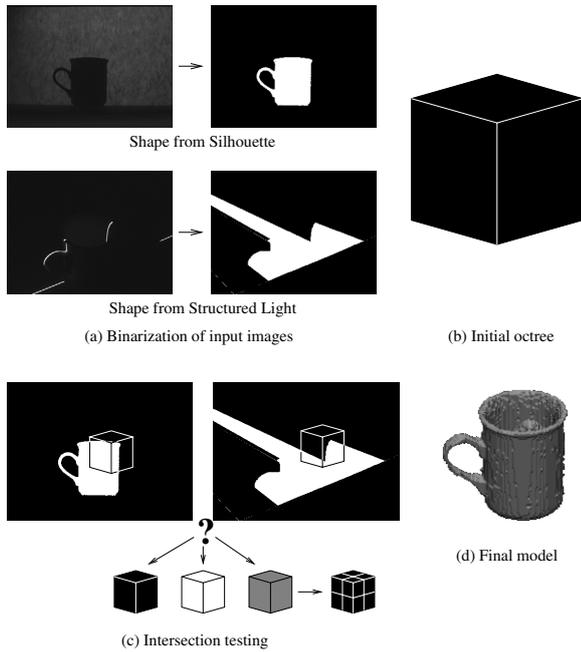


Figure 4. Algorithm overview.

An input image for SfL contains the projection of a laser plane onto the object (Figure 4a, 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. Based on the known position of the laser, an input image (Figure 4a, lower left image) is converted to an image approximating the intersection of the laser plane with the whole object (Figure 4a, lower right image).

Our approach builds a 3d model of an object performing the following steps (illustrated in Figure 4): First, both of the input images (SfS and SfL) are binarized such that the white image pixels *possibly* belong to the object and the black pixels *for sure* belong to the background (Figure 4a). Then, the initial octree containing one single root node marked "black" is built (Figure 4b). Black nodes are subsequently checked by projecting the nodes into all SfS binarized input images and intersecting them with the image silhouettes of the object (Figure 4c). As the result of the intersection the node can remain "black" (if it lies within the object) or set to "white" (it lies outside the object) or "grey" (it lies partly within and partly outside the object). If the resulting node is not white, it is projected into the binarized SfL 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 4d).

4. Results

For tests with synthetic objects we can build a model of a virtual camera and laser and create input images in a way that the images fit perfectly into the camera model. We assume having a virtual camera with focal length $f \approx 2 \text{ mm}$, placed on the y axis of the world coordinate system, 2000 mm away from its origin. We set the distance between two sensor elements of the camera to $d_x = d_y = 0.01 \text{ mm}$. The laser is located on the z axis of the world coordinate system, 850 mm away from its origin, and the turntable 250 mm below the x - y plane of the world coordinate system, with its rotational axis identical to the z world axis. We build input images with size 640×480 pixels, in which 1 pixel corresponds to 1 mm in the x - z plane of the world coordinate system. As the synthetic object we create a sphere with radius $r \approx 200 \text{ mm}$. Since the sphere does not contain any cavities, SfS can also reconstruct it completely. Therefore, we can measure the accuracy of each of the methods independently, as well as of the combined method.

In a test we build models using 360 views with the constant angle of 1° between two views, while increasing octree resolution. It turned out that the SfS method performed best with an octree resolution of 128^3 , where the approximation error was $+0.83\%$ of the actual volume, the structured light method with a resolution of 256^3 and $+0.29\%$ error (the other method produced there an error of -1.42%). In the second test we build models with constant octree resolution of 256^3 and increasing number of views. Regarding the number of views, there was no significant difference between the two methods. Using 20 instead of 360 views was sufficient for both methods to create models less than 1% different from the models built using 360 views.

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 5 shows the objects and their models built using 360 views for each of the underlying methods and the octree resolution 256^3 .

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 biggest 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.

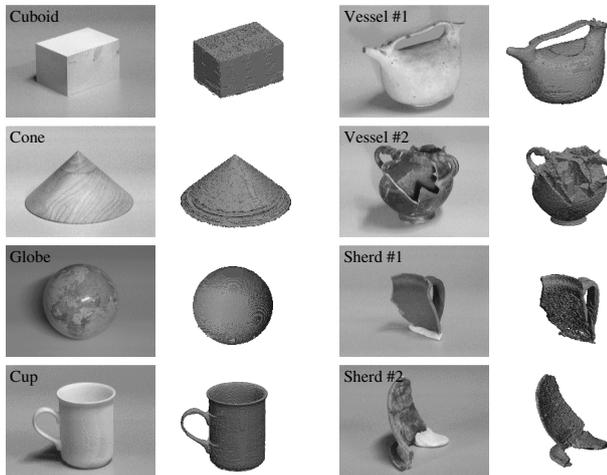


Figure 5. Real objects and their models.

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

In this paper a combination of a SfS method with a SfL method was presented, which creates a 3d model of an object from images of the object taken from different view-

points. The algorithm employs only simple matrix operations for all the transformations and it is fast, because even for highly detailed objects, a high resolution octree (256³ voxels) and a high number of input views (36), the computational time hardly exceeds 1 minute on a Pentium II. Already for a smaller number of views (12) the constructed models were very similar to the ones constructed from 36 views and they took less than 25 seconds of computational time.

For archaeological applications, the object surface has to be smoothed in order to be applicable to ceramic documentation, for classification, however, the accuracy of the method presented is sufficient since the projection of the decoration can be calculated and the volume estimation is much more precise than the estimated volume performed by archaeologists.

References

- [1] H. Baker. Three-dimensional modelling. In *Proc. of 5th Intl. Conf. on AI*, pages 649–655, 1977.
- [2] B. Baumgart. *Geometric Modeling for Computer Vision*. PhD thesis, Stanford AI, 1974.
- [3] P. Besl. Active, optical range imaging sensors. *MVA*, 1(2):127–152, 1988.
- [4] F. DePiero and M. Trivedi. 3-D computer vision using structured light: Design, calibration, and implementation issues. *Advances in Computers*, 43:243–278, 1996.
- [5] R. M. Haralick and L. G. Shapiro. Glossary of computer vision terms. *PR*, 24(1):69–93, 1991.
- [6] M. Kampel and R. Sablatnig. Automated 3d recording of archaeological pottery. In D. Bearman and F. Garzotto, editors, *Proc. of Intl. Conf. on Cultural Heritage and Technologies in the 3rd Millennium*, pages 169–182, 2001.
- [7] K. Kutulakos and S. Seitz. A theory of shape by space carving. *IJCV*, 38(3):197–216, 2000.
- [8] A. Laurentini. The visual hull concept for silhouette-based image understanding. *PAMI*, 16(2):150–162, 1994.
- [9] C. Liska and R. Sablatnig. Estimating the next sensor position based on surface characteristics. In *ICPR00*, volume I, pages 538–541, 2000.
- [10] W. N. Martin and J. K. Aggarwal. Volumetric description of objects from multiple views. *PAMI*, 5(2):150–158, 1983.
- [11] W. Niem. Error analysis for silhouette-based 3D shape estimation from multiple views. In N. Sarris and M. Strintzis, editors, *Proc. of Intl. Workshop on Synthetic-Natural Hybrid Coding and 3D Imaging*, pages 143–146, 1997.
- [12] M. Potmesil. Generating octree models of 3D objects from their silhouettes in a sequence of images. *CVGIP*, 40:1–29, 1987.
- [13] R. Szeliski. Rapid octree construction from image sequences. *CVGIP: Image Understanding*, 58(1):23–32, 1993.
- [14] S. Tosovic and R. Sablatnig. 3d modeling of archaeological vessels using shape from silhouette. In *Proc. of Conf. on 3-D Digital Imaging and Modeling*, pages 51–58, 2001.