Object Categorization

Axel Pinz
Graz University of Technology
Austria
axel.pinz@tugraz.at

Full text available at: http://dx.doi.org/10.1561/0600000003
Editorial Scope

Foundations and Trends® in Computer Graphics and Vision will publish survey and tutorial articles in the following topics:

- Rendering: Lighting models; Forward rendering; Inverse rendering; Image-based rendering; Non-photorealistic rendering; Graphics hardware; Visibility computation
- Shape: Surface reconstruction; Range imaging; Geometric modelling; Parameterization;
- Mesh simplification
- Animation: Motion capture and processing; Physics-based modelling; Character animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape Representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and Video Retrieval
- Video analysis and event recognition
- Medical Image Analysis
- Robot Localization and Navigation

Information for Librarians
Foundations and Trends® in Computer Graphics and Vision, 2005, Volume 1, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.
Object Categorization

Axel Pinz

1 Graz University of Technology, Austria, axel.pinz@tugraz.at

Abstract

This article presents foundations, original research and trends in the field of object categorization by computer vision methods. The research goals in object categorization are to detect objects in images and to determine the object’s categories. Categorization aims for the recognition of generic classes of objects, and thus has also been termed ‘generic object recognition’. This is in contrast to the recognition of specific, individual objects. While humans are usually better in generic than in specific recognition, categorization is much harder to achieve for today’s computer architectures and algorithms. Major problems are related to the concept of a ‘visual category’, where a successful recognition algorithm has to manage large intra-class variabilities versus sometimes marginal inter-class differences. It turns out that several techniques which are useful for specific recognition can also be adapted to categorization, but there are also a number of recent developments in learning, representation and detection that are especially tailored to categorization.

Recent results have established various categorization methods that are based on local salient structures in the images. Some of these methods use just a ‘bag of keypoints’ model. Others include a certain amount of geometric modeling of 2D spatial relations between parts, or ‘constellations’ of parts. There is now a certain maturity in these approaches...
and they achieve excellent recognition results on rather complex image databases. Further work focused on the description of shape and object contour for categorization is only just emerging. However, there remain a number of important open questions, which also define current and future research directions. These issues include localization abilities, required supervision, the handling of many categories, online and incremental learning, and the use of a ‘visual alphabet’, to name a few. These aspects are illustrated by the discussion of several current approaches, including our own patch-based system and our boundary fragment-model. The article closes with a summary and a discussion of promising future research directions.
Contents

1 Introduction 1

1.1 Problem statement 2
1.2 Historical development 3
1.3 Potential applications 5
1.4 Outline of this review article 5

2 Categorization as an Issue of ... 7

2.1 ... classification 7
2.2 ... learning 10
2.3 ... representation 12
2.4 ... localization 18
2.5 ... datasets 21
2.6 ... evaluation 24
2.7 ... system integration 27

3 Building Blocks for Categorization 29

3.1 Scale in space and time 29
3.2 Saliency, key-features, points and regions of interest 36
3.3 Object models 49
3.4 Learning and recognition 62
This article provides a review of existing representations, algorithms, systems and databases for visual object categorization. It describes the state of the art in this field, which has been a long standing goal, and is still a mainly unsolved problem in computer vision research. The time chosen for writing is motivated by recent success in recognition from local, salient parts, which can be considered a significant step towards object categorization.

Who are the supposed readers of this document, and what potential benefits are there for them? Students and graduate students in computer vision will get a thorough review of the state of the art in visual object categorization. Researchers in computer vision might benefit from a more complete point of view, including a number of approaches which they have not focused on within the scope of their own research. Researchers in related fields should find this article a valuable reference.

But the article goes beyond a pure review of the state of the art. It includes original research in categorization, presents a prototype system for categorization, discusses our databases and provides experimental results on object categorization and localization in still images.
1.1 Problem statement

We can define visual object categorization as the process of assigning a specific object to a certain category. This process has also been termed ‘generic object recognition’ (generic OR), and it is in contrast to ‘specific OR’, which deals with the recognition of a specific, individual object. Examples of categories in generic OR are people, children, dogs, cars, bikes or dishes, while specific OR might aim at recognizing a certain individual, like Albert Einstein, or a specific object like my car. An individual object might also be termed a specific instance of a more generic category. Categories can also be organized in hierarchies (child – human being – mammal), and categories might overlap – a tall glass might be used as a vase. Throughout the remainder of this article, we will use the terms ‘categorization’ for visual object categorization or generic OR, and ‘specific OR’ for the recognition of individual objects.

Looking at humans, and comparing their recognition performance with artificial systems, it turns out that humans are much better in categorization than machines, but specific OR can often be handled more efficiently, reliably or simply faster by an artificial vision system. VanRullen and Thorpe [199] point out that humans can perform ultra-rapid categorization tasks. They can decide whether a briefly flashed image belongs to a certain category in less than 150ms, and they provide experimental evidence for the two categories ‘animal’ and ‘means of transport’. On the other hand, there are numerous solutions to industrial inspection, which recognize and localize specific objects much faster and much more reliably than humans can do (see [31] for an example of such an industrial product, and [76] for the underlying theoretical foundations). A further aspect of categorization is the sheer number of visual categories. There is evidence from cognitive psychology that humans deal with about 30,000 different categories (see Biederman [21]). This would require solving currently intractable computational complexity.

This article sets out to answer the following questions: How can artificial systems perform categorization? What are the key building blocks that are required to build a categorization system? What are the main challenges? What are the bottlenecks and unsolved problems? This will
also shed light on the more general question: Why is categorization simpler for humans than for machines and why is specific OR simpler for machines than for humans?

1.2 Historical development

In the following, I give a very brief sketch of some major landmarks in the history of object recognition research. This is not meant to be a complete review, but rather some useful information to pave the way for later discussion. One of the major early landmarks is certainly the work of David Marr [126], who proposed viewer-centered and object-centered representational levels (image – primal sketch – 2-1/2D sketch – 3D object model), as well as visual modules which can be used to generate these descriptions (e.g. ‘shape from X’ to produce a 2-1/2D sketch). Marr’s ideas influenced at least a decade of research, and have led to a so-called ‘reconstruction school’ which advocates that 3D reconstruction and 3D modeling of a scene (and thus of the objects in the scene) are necessary for further reasoning.

On the other hand, there is the ‘recognition school’ which favors working in the 2D domain, with 2D images, features and descriptors which are extracted from these images. Their pattern classification [43] or pattern recognition [145] approach is fundamentally different from the reconstructionist paradigm. Much of the content of this article actually is in the spirit of a ‘recognition school’ approach. We will discuss, for example, the ‘bag of keypoints’ approach, in which salient points are extracted from images, and descriptors are calculated to form feature vectors. These feature vectors can be used to learn a discriminative model from training images, and to recognize (categorize) test images. But we will also present the generative ‘constellation’ model, which employs a ‘light’ 2D geometry in terms of spatial 2D relations between key parts of the object model.

Only very recently have we seen efforts to combine discriminative and generative approaches in categorization research. This confluence of recognition and reconstruction schools has already been predicted by Aloimonos and Shulman [4] in 1989.
There are further milestones which should be mentioned. Biederman [20, 21] proposed his ‘recognition by components’ (RBC) theory. Volumetric primitives, so-called ‘geons’ can be used to recognize objects in a qualitative (and thus generic) way. While this theory is quite elegant, its implementations (see [18, 39]) lacked due to low level vision problems, so that geon-based recognition has not been applied to real-world categorization problems. Research in perceptual grouping proceeds in a similar manner [167]. Low-level 2D primitives are grouped to build object descriptions, either in a pure bottom-up (data driven) manner, or top-down, including prior knowledge (models) about the expected image content. At the other end of the spectrum of potential solutions to OR, we find the idea of purely image- or ‘appearance’-based recognition, for instance in parametric eigenspace [143]. This idea has triggered a vast number of extremely successful appearance-based approaches to specific OR.

In general, there has been more research in specific OR than in categorization in the past. Success in specific OR has influenced a number of approaches to categorization, although most of the developed algorithms for specific OR are not directly applicable to categorization. There is a paradigm of specific OR by alignment, in which spatial correspondence between groups of image features and model features is found by searching for the geometric transformation that aligns these features best. This includes affine transformations for planar objects [51] and 3D model to 2D image feature matching [121]. Another way to compare image and model features is to extract features which are invariant against certain geometric [122] or radiometric distortions [13]. Efficient indexing is needed, when a database of potentially many object models has to be matched against features extracted from a query image that contains a certain specific object. This can, for instance, be done by geometric hashing [207], a technique which is robust against partial occlusion and geometric transformations. Finally, the success in global appearance-based recognition [143] has motivated research in local appearance-based methods for specific OR [122]. At this point, we can observe that techniques for specific OR and for categorization meet.
1.3. Potential applications

In categorization, these various ideas have led to the development of a number of recent approaches which try to:

- model appearance more locally,
- group simple geometric primitives, and
- use learning algorithms to find common patterns that can be shared over many individuals of a category.

Within the past 5 years, we have seen a rapid development and rise in the success of object categorization in increasingly difficult, cluttered, and realistic scenes.

We can also observe a number of contributions from related fields as machine learning, neurosciences and cognitive psychology.

1.3 Potential applications

There are a number of obvious applications of categorization to image database annotation, image retrieval and video annotation. But potential applications of categorization go far beyond that. Reliable categorization in real-time will open up applications in surveillance, driver assistance, autonomous robots, interactive games, virtual and augmented reality and telecommunications. A more general view might include systems for ‘cognitive personal assistance’ with many potential aspects, ranging from user support in complex environments to very basic support capabilities for elderly or disabled people.

1.4 Outline of this review article

The article is structured in three major parts (Sections 2–4). I start with an in-depth analysis of major issues related to solving the problem of categorization mentioned in Section 2. This analysis provides at the same time an introduction to the main topics, which are then discussed in detail in Section 3 which presents the major building blocks for

---

1 This may partly be related to recent European research initiatives. There has been substantial funding of basic research in ‘Cognitive Vision’ within the 5th framework program of the European Union, with an even broader perspective of ‘Cognitive Systems’ in the current, 6th framework program. There has been strong support of categorization research within these programs.
Introduction

categorization systems. Finally, Section 4 presents two major aspects of our own research in categorization: a region-based approach, and categorization with a boundary-fragment-model.

You will probably recognize that the subject is quite broad and heterogeneous (ranging from the representation of scale in images, over machine learning, to 2D spatial models for categorization). Thus, there is no isolated section on the ‘state of the art’ and related work. I prefer, rather, to cite relevant publications throughout the whole article, which is hopefully more useful to you, the potential reader.

Finally, there is a common thread, which should provide some extra value for those who manage to read the complete article sequentially. However, many sections stand on their own and may also be consulted individually.
References


References

References

References


References

References


References

References


References

References
