名称及び氏名	博士（工学）	Martin Klinkigt
学位授与の日付	平成24年9月30日	
論文名	「Semantic Retrieval of Images Using Object Recognition Systems by Learning from Internet Communities」 （インターネットコミュニティからの学習による物体認識システムを用いた画像の意味的検索）」	
論文審査委員	主査	黄瀬　浩一
	副査	辻　洋
	副査	市橋　秀友
	副査	Andreas Dengel

論文要旨

With the availability of cheap image photography the number of images in private collections increases rapidly. Digital photography pushed this trend even further and several hundred of images taken even in a short trip are quite common. However, in the past and now the question which always arises is how these images can be organized to retrieve them in an efficient manner.

One possible solution is to use semantics to describe the content of images. For example by naming “Holiday 2011”, “Paris” and “Ben” the user would retrieve all images taken during the holiday trip 2011 in Paris showing his friend Ben. Applications like Google’s Picasa or Apple’s iPhoto enable the user such a semantic description. But why such systems fail to become widely used? Often the designers of such systems forget about the fact that time-saving is the most important reason to use a system. The problem for the previously proposed systems is that the user first has to provide all this information for all possible search criteria. Keeping in mind that such image collections grow rather fast it would result in much handwork done by the user.

To address this major burden in semantic retrieval of images we propose to use computer vision. In computer vision a large number of systems accepted the challenge in providing such information. This task is called object recognition and can be separated into specific and generic. While the task of specific object recognition is to name the instance of an object as for example “Toyota Prius”, the task of generic object recognition is to name a class of the object for example “car”. Many recently proposed systems achieve remarkable recognition performance leading to the hypothesis that this problem has been solved. However, this is not the case since this recognition performance is often only achieved at high costs. Most of the constraints applied by existing systems are only acceptable in professional environments. Our focus is to support non-professional user and, therefore, generically applicable systems are required.

To achieve this goal our system first analyzes the data stored on the user’s computer to detect his interest. Having this interest the system utilizes Internet communities to learn visual information in a fully autonomous manner. Previously proposed systems normally request the user to teach the system. The motivation to use such communities is their reliability.

After our system has finished its learning it provides specific and generic object recognition. Specific object recognition provides information for example about a certain building like the Eifel Tower in Paris or the Toyota Prius shown in the image. Being more specific as for example distinguish among different Prius is not the focus of our system. The visual differences between different Prius are not significant enough to make a reliable decision for one of them. Since the information received by Internet communities is far from being complete the generic object recognition system provides information in the case where no learning example was provided. These are of abstract level as for example car.

By using such object recognition systems the images of the user are semantically enriched. The user can navigate in his image collection and find the desired images by naming their content, which reduces the search time as compared to manually checking all images.

In detail the thesis is arranged into the following chapters:

In Chapter 1 we explain the problem of organizing images. We give a short overview of our solution how object recognition can address this problem.

In Chapter 2 we put the focus on semantic aspects which we propose to address the problem of organizing images. Furthermore, we lay the background needed for this work and show how we provide an image training dataset for our system by using information available from the Internet. In our system we utilize Wikipedia as a source for training data which is more reliable as compared to other sources as for example photo sharing communities.

In Chapter 3 we give the basic concepts of object recognition and explain our solution to recognize specific objects as for example “Toyota Prius”. We explain the importance of local features and how we improve the recognition performance by 3% with the help of a shape context. With this shape context the system analyzes where the surrounding features are located.

In Chapter 4 we point out the problems of the shape context recognizing partially visible objects and present our solution by working with a reference point. While the shape context is suited for local shapes, we show that the reference point is effective to express the global shape. With the help of this reference point we are able to recognize even partially covered objects as we show with evaluations on several datasets. With several experiments we confirm an improved recognition performance of 6% in terms of mean average precision compared to other shape models designed for this recognition problem.

In Chapter 5 we combine the proposed solutions in chapters 3 and 4 together with a preprocessing to a stable system addressing various problems by working with images. During this preprocessing the system detects image patches defined over regions of similar color and uses them to clean the image from ambiguous features. With this preprocessing we achieve stable recognition performance even with artificially increased databases.

In Chapter 6 we address the remaining problematic cases for specific object recognition, e.g., if the object is not known to the system. We propose to go over to generic object recognition to name the category of an object. The highest burden for generic object recognition is the preparation of the model description which is normally done by the user. In our system this task is again pushed to the Internet community which provides all needed information. In evaluations we show that our system can return the correct category in 50% of the test images, while the specific object can only be found for 14% of the test images.

In Chapter 7 we finally conclude this work by summarizing the previous chapters and proposing possible solutions for the open problems.

審査結果の要旨

本論文は、学習機構を備えた物体認識システムを適用して画像の意味的検索を実現する

研究の成果をまとめたものであり、インターネットコミュニティを学習の情報源として用いる点に特徴がある。成果は以下のようにまとめられる。

（１）　物体認識の辞書を学習するための情報源として、インターネットコミュニティで作成されたWikipediaなどに着目する重要性を、他の情報源との比較に基づいて指摘した。また、意味的検索のためには、物体のインスタンスを認識するための特定物体認識と、物体のクラスを認識するための一般物体認識の双方が必要であることを述べた。

（２）　特定物体認識の手法として、形状文脈を用いる手法を提案するとともに、新たに考案した参照点に基づく照合方法を同時に適用することで、さらに精度が向上することを実験的に検証した。

（３）　複雑背景下での特定物体認識の精度を向上させるため、画像を均質な小領域に分割し、その後、各小領域に対して特定物体認識手法を適用する方式を提案するとともに、その性能を実験によって検証した。

（４）　特定物体認識を適用するためには既に認識システムがその物体を知っている必要がある。この条件が成り立たない場合には、一般物体認識を適用する必要がある。一般物体認識のための辞書を学習するための情報源として、同様にインターネットコミュニティで作成されたWikipediaなどを用いる手法を提案した。特徴は、物体のクラスを直接定義するのではなく、属性を用いて定義し、属性と画像特徴との関係を学習する点にある。これにより、直接定義する場合に比べて、より少ないデータで適用範囲の広い結果を学習することが可能となった。

以上の諸成果は、物体認識を用いて画像の意味的検索を行う際に最も大きな問題となる「学習」に対して新たな方法を与えるものであり、画像検索に貢献するところ大である。また、申請者が自立して研究活動を行うのに必要な能力と学識を有することを証したものである。