

論文の内容の要旨

論文題目 A Real-Time Object Tracking System With On-Line Feature Learning

(オンライン形状学習機能を備えた実時間物体追跡システム)

氏 名 趙 普 社

Object tracking plays an important role in many applications, such as video surveillance, human-computer interface, vehicle navigation, and robot control. It is generally defined as a problem of estimating the position of an object over a sequence of images. In practical applications, there are many factors that make this problem complex, such as illumination variation, appearance change, shape deformation, partial occlusion, and camera motion. Moreover, lots of applications require real-time responses. Therefore, the development of real-time working systems is of essential importance. In order to accomplish such a challenging task, a number of tracking algorithms and real-time working systems have been developed in recent two decades. However, it is still difficult to compete with the tracking ability of human brain.

In this thesis, we proposed a solution to the challenging tracking task based on the consideration of efficient implementation as priority, and realized functions as follows. At first, a hardware-friendly tracking framework is designed and implemented on FPGA, which is compatible with VLSI technology. This framework, named multiple candidate regeneration (MCR), is developed as a simple but high-speed and efficient searching algorithm. The basic idea was inherited from the particle filter (PF), but the algorithm has been greatly modified and simplified from the original PF so that it can be implemented in VLSI hardware very efficiently. Because of its simplicity, MCR's weighting and updating processes can probably be explained from the viewpoint of PF. However, there are still several significant differences between them, because the MCR is developed by simplifying the visual tracking problem itself, aiming to improve the accuracy based on simple hardware implementation. For example, the number of candidates in the MCR is not strictly depended on by the performance as in the PF, because in tracking task, we can make the assumption that if the hardware is fast enough, the differences (location, appearance) between two consecutive frames of images should be sufficiently small, which means much less templates are needed. In the development process, several problems, which limited the hardware performance, have been resolved, such as complex computation, data transmission and cost of hardware resources. The proposed architecture of the MCR achieved 150 frames per second (f/s) on FPGA, and can reach about 900 f/s if it is implemented on VLSI with on-chip image sensor. This solution owns several advantages. First, it can work at high frame rate

if necessary, which can simplify the work of localization. It can meet the requirement of higher processing speed in some complex intelligence systems, which seems difficult to achieve by conventional solutions. Second, the system can be extended to use in many applications because of its flexibility. Third, since the processing speed is much faster than the frame rate, it is probable that the ability of this system is improved to be more accurate and robust. The system was implemented on a Terasic DE3 FPGA board. Under the operating frequency of 60 MHz, the experimental system achieved a processing ability of 0.8 ms per frame in tracking a 64×64-pixel size object image in 640×480-pixel size video sequences.

In tracking algorithms, how to represent the target image is of particular importance because it greatly influences the tracking performance under certain tracking framework. Color, edge, and texture are typical attributes used for representing objects. A number of other features, including active contour, scale-invariant feature transform (SIFT) feature, oriented energy, and optical flow, are also used in many works. Some works also combine these features or incorporate on-line learning of the appearance model of an object and background. In this thesis, we have aimed to establish both robustness of object representation and the real-time performance of the processing, because feature extraction is usually a time consuming process. It was well known that the visual perception of animals relies heavily on the directional edges. In this work, therefore, the directional-edge-based image feature representation algorithm is employed to represent the object image. Robust performance of the directional-edge-based algorithms has already been demonstrated in various image recognition applications. In addition, dedicated VLSI chips for efficient directional edge detection and image vector generation have also been developed for object recognition systems.

Whether a tracking system can be easily extended for various purposes is also very important. This thesis contains a detailed discussion on extending the function of the system, including hardware implementation on VLSI, multiple-object tracking, full-occlusion and initialization problems, and employing of state vector. The architecture of this system is compatible with VLSI design, and may reach better performance on VLSI. For the multiple-object tracking, an efficient method is proposed to allocate the limited hardware resources. For the full-occlusion and initialization problems, a searching algorithm based on proposed system is developed. By using the state vector, more attributes can be estimated for achieving more information about the object. In this part, the attributes of the object are combined and transformed to a high-dimension vector, which replaces the two-dimension location vector (x , y). It also improves the tracking ability and accuracy as reflected in experiments.

The following parts of the thesis are focused on the learning ability of the system. For object tracking, one promising direction is to consider the object tracking as a binary classification problem, and employ discriminative methods in the tracking framework. Support vector machine (SVM), as a powerful

classification scheme, has been used in many tracking algorithms, benefiting the algorithms with accurate localization and flexible modeling of the target. The SVM works as an appearance model of the target by changing its boundary while training. One feature of SVM is that the boundary is represented by the combination of support vectors, and the number of support vectors is usually a small portion of the total training dataset. This feature becomes very important when implementing the tracking algorithm on hardware, because the hardware resources always have a limitation.

Despite the good performance of these learning-based algorithms, they suffer from several practical problems. Some work builds a superior SVM classifier and gives good results in tracking vehicles. However, the off-line training mechanism employed in the work requires a large number of training samples selected manually and does not support updating the training examples. In some research, all samples learned from each frame of an image sequence are stored for training the SVM. This causes a large memory cost if it is used in a long-duration task. In some work, a simple strategy is employed to determine new training samples, which may cause “drift problem”. Moreover, these algorithms do not consider their real-time performances, which is in fact of great importance in object tracking applications. This is mainly because of the complex training process of SVM. Especially for the on-line learning SVMs, frequently repeated training and predictions make this problem even worse. Therefore, in order to extend the power of SVM in most of the general tracking applications, it is necessary to develop a proper tracking framework and a VLSI hardware-implementation friendly structure for the SVM-based algorithm.

In this thesis, based on the MCR tracking system, a real-time visual tracking algorithm is presented employing an on-line support vector machine (SVM) scheme. A novel training framework is proposed, which enables us to select reliable training examples from the image sequence for tracking. The tracking framework includes how to update training examples and how to select test samples and make prediction of the target location. Different from other algorithms, this framework gives a rule guiding the selection of target training samples. When the target changes its appearance significantly, the system may fail to localize the target because the classifier misclassifies the target image to the background image category. In order to solve this problem, background samples are utilized to predict the location of the target image. Unlike the moving target image, most of the background sample images are stable. As a result, high-accuracy tracking has been established. In addition, regarding the selection of target examples for on-line training of SVM, a new selection rule has been introduced.

Multiple candidate regeneration is also employed to decrease the computational cost, and the directional-edge-based feature representation algorithm is used to represent images robustly as well as compactly. The structure of the algorithm is designed especially for real-time performance, which can extend the advantages of SVM to most of the general tracking applications. The algorithm has been evaluated on challenging video sequences

and showed robust tracking ability with accurate tracking results. The hardware implementation is also discussed, while verification has been done to prove the real-time ability of this algorithm.

The on-line SVM learning requires repeated training and predicting. The predicting process always contains computation of thousands of test samples in conventional algorithms, preventing these algorithms from working in real-time. In this process, not only the SVM, but also the feature extraction of each sample will cost lots of time. Based on a SVM chip developed in our group, the most complex part in this algorithm can be computed efficiently. At the same time, multiple candidate regeneration is employed to reduce the computational cost without sacrificing the tracking accuracy. In addition, the directional-edge-feature vector representation, whose VLSI implementation has been proposed in, is employed to represent the sample images. By using this hardware-friendly structure, real-time tracking ability can be achieved. The hardware architecture for realizing this kind of real-time tracking system is discussed in detail.

From the SVM-based system, several parameters are observed and evaluated. One the most important parameters is the number of the support vectors (related to the number of training examples). This value of this parameter is around 40 to 50 in the tracking process of one video clip. Therefore, we can draw the conclusion that much less templates are needed in a tracking task to describe the appearance of the object than in an object recognition task. Based on this analysis, a simpler classifier was considered to use in the system for classifying object sub-images from other sub-images. As a result, the nearest neighbor (NN) classifier is employed. Different from other researches, the implementation of NN classifier in the proposed system is weak as a normal classifier but sufficient for the requirement in the tracking system. Several comparison experiments were carried out to verify the effectiveness. After optimization, the maximum numbers of the templates stored for object class and non-object class were both set to 50, and experiments on object tracking database were carried out. The experimental results illustrated that the tracking accuracy was comparative to the SVM-based tracking system, while it was much simpler from the viewpoint of hardware implementation.

In summary, this thesis presents a real-time solution to object tracking task with on-line learning capability. The robust feature learning capability of the system is realized by employing the SVM and NN classifiers in the tracking problem. And a new tracking framework for the classifier-based algorithm is also designed. The hardware implementation problem is carefully considered from the most beginning. Then hardware-friendly algorithm and architecture were designed and a real-time tracking system has been implemented. Extensive experiments were performed for evaluation on the tracking system. The thesis also contains detailed and extensive discussions about improvement on this system.