論文の要旨
（Thesis Summary）

氏名　　　駱 愛文
（Name）　Aiwen Luo

論文題目(Thesis Title)

Research on Efficient Vision-Based Hardware Architecture for Feature Representation and Object Detection

(特徴表現及び物体検出のための効率的なビジョンベースのハードウェアアーキテクチャに関する研究)

Machine learning (ML) is described by Arthur Samuel as the field of study that gives computers the ability to learn without being explicitly programmed. The methods for solving ML problem can be classified into two groups, i.e. supervised learning and unsupervised learning. Pattern recognition is a branch of ML that focuses on the recognition of patterns and regularities in data. Pattern recognition systems can be trained from labeled training data in supervised learning, but also can be used to discover previously unknown patterns in unsupervised learning when no labeled data are available.

Feature representation is becoming an important prerequisite for building ML models to achieve high detection performance, in particular in vision-based applications. A wealth of feature-extraction techniques are being presented in this research field. To meet for the embedded systems with limited resources, this thesis developed a simplified scale-invariant speeded up robust features (SURF) approach, which can decrease object-matching complexity and enhance computational performance significantly, to extract the feature vectors in the image. Whereas, the histogram of oriented gradient (HOG) is a widely accepted feature for object detection, which attains high accuracy against changes of illumination of variously textured objects.

Comparing to hardware solutions, the software technologies for feature extraction are more affected by computationally demanding algorithms so that the power dissipation of software technologies is burdensome for mobile applications and the processing speed becomes a bottleneck in case of real-time (>30 frames/s) object detection. Consequently, the achievable advantages of faster processing speed and lower power consumption lead to application-specific integrated circuit (ASIC) solutions.

Therefore, a resource-efficient coprocessor for a simplified SURF descriptor, employing Haar-like wavelets as feature vectors (FVs), is developed and a prototype is fabricated in 180nm CMOS technology in this research. A pipelined hardware architecture with the low computational budget, that directly uses the serially-inputted pixel data without pre-storage, is

employed for the local image cells. With a novel multiplexing process of the cell-based partial feature vectors, developed in this research work, the multidimensional feature vectors of multiple scan windows can be extracted simultaneously. The high affordable frame rate (up to 325 frames/s) meets for real-time processing while only 12 k bytes on-chip memory space consumes for feature extraction with the developed simplified SURF descriptor in this research. The die area of the prototype is reduced by more than 98.25% comparing to previous works. The normalized power consumption is 8.723 mW with a constant field scaling to 65 nm CMOS, which is 96.04% less power dissipation than previous works and is much more suitable for mobile applications with limited battery capacity.

Afterwards, the cell-based local FVs are sequentially outputted for the parallelized window-level multidimensional feature representation, and subsequent object matching directly without normalization. In this research, the concept of "regular rule of reusing times" (RRRT) of each cell is proposed so that the multiple window-based descriptor vectors can be constructed in parallel with high speed through access to look-up tables.

The first object detection framework, which combines with the pipelined feature extraction, the cell-based sliding window algorithm and the nearest neighbor search (NNS) classifier is verified in 65 nm CMOS technology. During the entire object detection procedure, the prototype coprocessor can achieve 34.6 fps VGA (640×480) frame rate when working at 200 MHz frequency. Furthermore, the requirements of 1.26 mm^2 die area and 26 k bytes on-chip memory result in 31.49 mW power dissipation at 1.0V supply voltage. The coprocessor can substantially reduce computation cost without significant degradation of classification accuracy. The maximal detection accuracy for front car can achieve 98.78% while 93.90% for pedestrian detection.

Since the object detection is greatly hampered by a large amount of high-dimensional FVs in the former-proposed feature construction scheme, I further employ partial least squares (PLS) analysis to project the FVs data onto a much lower dimensional subspace, to reduce the computational amount and to save significantly in resource consumption. Up to 242.48 MHz frequency of the developed hardware architecture is verified for pedestrian detection operated on the Altera Stratix-IV field-programmable gate array (FPGA) platform.

Furthermore, a block-based sliding window algorithm with normalized feature vectors by L1-norm scheme is developed for the second object detection framework that combines with the HOG descriptor and support vector machine (SVM) classifier. The improved detection framework aims to increase the pedestrian detection accuracy and robustness. Since fixed application-field parameters will normally make a hardware architecture unsuitable for practical scenarios, flexible regulation of image size, cell size (CS), etc., is desired in the second object detection framework to meet for a variety of mobile applications with the usually occurring changes under various circumstances. Variable image resolutions up to 1024 (width) × ∞ (height) pixels with the same hardware architecture without any modification can be handled by the reported research results. Five-level CSs, i.e., 2×2, 4×4, 8×8, 16×16, 32×32 pixels with an unlimited image width, applied in the on-chip FV-extraction circuit, are implemented in another 65nm CMOS prototype ASIC. Up to 125 fps of XGA (1024×768) images can be detected by the developed coprocessor. 41 k bytes on-chip memory and 2.86 mm^2 die area are consumed in this scheme. The on-chip memory in this scheme seems to increase comparing to the first framework with the NNS classifier due to the higher dimensionality of the HOG descriptor (3780 dimensions) than SURF (1680 dimensions). Moreover, the on-chip memory requirement increases multiply with the reference number for the NNS classifier, whereas the SVM coefficients number is fixed with a given window size. Thus the on-chip storage and area requirements for the second framework can be smaller when cooperating with more references. The power consumes 21.3 mW when the coprocessor operates at 125 MHz frequency and 1.0 V supply voltage.

In conclusion, this research developed a cell-based feature extraction architecture and two object detection frameworks combined with NNS and SVM respectively. Both the NNS and the SVM classifiers are applied for verifying the high detection performance of the proposed hardware architectures, which are verified to meet the requirements of mobile applications.