Hardware Optimized Image Processing Algorithms for Industrial High Speed Applications

JÖRG VELTEN, MARCO KRIPS, ANTON KUMMERT
Department of Electrical and Information Engineering
University of Wuppertal
42097 Wuppertal
GERMANY

Abstract: - Industrial high speed image processing is a challenging field of application. Conflictive requirements concerning computational power and economic expenses have to be met. The present paper describes the development of hardware optimized algorithms for implementation in video based industrial safety devices, where complex production processes have to be protected. Selection of algorithms was done with respect to an optimized implementability in FPGAs.

Key-Words: - image processing, video based safety device

1 Introduction

The usual choice of hardware for high speed image processing and recognition tasks are proprietary solutions with several specialized processors which are integrated in a parallel processing architecture. Implementation in industrial environments often fails because of very restricted objectives regarding device prices and technical abilities. For this reason, development of video based safety devices is still in its beginnings. Nevertheless, even today's technically restricted options can lead to significant improvements with respect to abilities of conventional protection devices. An often employed safety device for preventing accidents is the light curtain. It consists of several light barriers which are arranged in a column. The distance between a pair of light columns is small enough to detect any intruding object greater than a certain diameter. All light rows together define a planar area which can not be penetrated by these objects without being notified. The static form of this area is the reason for development of a more flexible device, based on video images. The latter can only be implemented by using a-priori knowledge with respect to the safeguarding situation, a high grade of adaptation to a certain workplace, and the use of efficient combinations of algorithms and hardware. The present paper describes the implementation of suitable combinations of algorithms and hardware for a video based protective device at press brakes. Safeguarding of different machines can be implemented following similar adaptation processes. The announced introduction of a-priori knowledge leads to the necessity of giving a short introduction to press brake components and typical workflows and dangerous situations.

2 Protective Devices at Press Brakes

Press brakes mainly consist of a punch (upper tool, movable) that presses the stock, usually prefabricated sheet metal in a die (lower tool). The movable tool is able to apply forces of up to 3200 kN and can be handled in operating speed (up to 10 mm/s) or in rapid speed (up to 250 mm/s) for opening and closing operations. Figure 1 shows the image of a typical press brake work place.

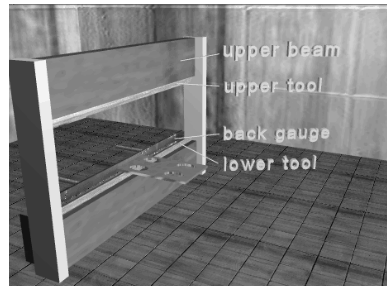

Figure 1: Image of a typical press brake

The usual workflow at press brakes comprises of

1. insertion of a work piece,
2. closing of the machine until hands do not fit between the tools anymore (usually in rapid speed)
3. bending (in operating speed) and
4. extraction of the work piece.

Large presses allow several different bends of one work piece to take place at different set-up stations of the same press. Safeguarding of such machines is nowadays often done by horizontal oriented light curtains, which establish a not unperceived penetrable security plane - the machine is stopped, if this plane is violated at any position. While many bending operations can be secured by these devices, problems arise from large work pieces and problematic shapes. Those parts have often to be bent using an inconvenient, e.g. two hand control, which prevents hands from being in danger area, when a machine cycle starts. Uncomfortable safety devices are unfortunately often bypassed or simply switched off, especially if the worker is doing piecework. Consequently, the design of a video based protective device, which avoids inconvenient processing steps is of special interest. A careful analysis of dangerous situations and its possible avoidance and detection is necessary, which leads to the observation that occurrence of possible dangerous situations differs in time and location. Phase 2 is extremely dangerous since even large extremities can reach between the tools, and closing of the machine often happens in rapid speed. Danger of cutting between upper and lower tool is reduced during phase 3, since the press is almost closed and driven in operating speed. However, during this phase, bending of boxes can cause bruises of hands between upper beam and work piece, i.e. at a different location. Nevertheless, a protection for all situations of danger is possible, if hands can be detected (validated) outside a predefined, but varying danger zone, i.e. different danger zones can be defined for each of the above-defined steps of the workflow. In case of a video based device, image acquisition and processing speed determine the required zone for avoidance of injuries. Because the follow-up time of usual press brakes is approximately 20 ms and the frame duration of usual cameras equals the same amount (at 50 Hz frame rate), 20 ms is also the designated processing time for a danger detection algorithm. The improvement over conventional protective devices results from a higher grade of flexibility, which is achieved by the fact that danger-zones can vary during a processing cycle.

3 Implementation Considerations

Requirements for high computational demands and a low cost implementation (magnitude of economic effort preferably as for light curtains) lead to the necessity of a careful adjustment of algorithms and hardware architecture. The latter can be obtained by keeping the following guidelines in mind: Early processing steps should be implemented in filter structures to handle high data rates. Algorithms, which require random data access, are placed at the end of the processing queue. Algorithms that reduce data word length lead to a reduction of computational effort and hardware complexity. It is furthermore convenient to start by extraction of local information, i.e. by means of small filter masks and apply regional or global operations to signals of already reduced data rate, thus, at the end of the processing queue. In view of the high computational requirements and the need for a freely definable processing architecture, ASIC technology would be the preferred choice, which leads to very high development costs. On the other hand the relatively small number of instances leads to the demand for use of standard components. FPGAs represent a trade-off between both requirements. The analysis of convenient algorithms is therefore performed under consideration of elementary FPGA components, i.e. logic elements (LE). The latter consist of a four to one mapping of Boolean signals and a latch for result storage, as shown in figure 2. Commercially available low

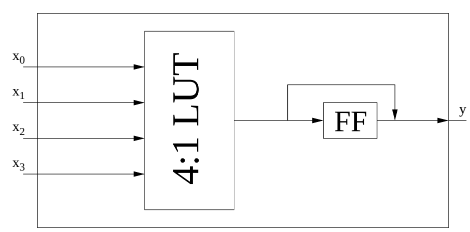

Figure 2: Schematic description of a logic element (LE)

cost FPGAs as the Xilinx XC2S-200 include 4704 programmable LEs of this kind. Because of the aforesaid reasons, it is convenient to also restrict the hardware set-up to only one color camera mounted at the top of the machine, which gives images from a bird's eye view perspective. Consequently, image recognition will be based on the usual R- G- and B- signals with 8-bit resolution each. Figure 3 shows the intensity information of an example image. The monitoring of operativeness is not considered in the present paper. Two independent algorithms, which follow this implementation concept, are described below. They follow different principles for ensuring an inviolated "region of danger". The first

Figure 3: Example image of a press brake (grayscale)

algorithm follows the principle of detecting any object intruding a buffer zone between worker and machine, while the second one verifies hands at risk-less positions.

4 Safety Area Surveillance

This approach is followed by conventional safety devices at press brakes. Light curtains consist of several light barriers, arranged in a row and implement thus a planar surveillance surface. Any penetrating object of a certain diameter is recognized, which effectively establishes a danger zone between the light curtain and the machine, if this zone is not large enough to completely conceal a worker. The emitted light signal is encoded for the purpose of protection against spoofing. In case of a video based detection system, similar results can be obtained calculating the difference of a camera image to a reference image of the empty work place. The actively emitted time coded signal of the anti-spoofing feature has to be imitated by a spatially coded signal, which can be implemented by using a predefined image background pattern. Interrelation between changes in illumination and the calculated difference image is reduced by using binary edge images for differentiation instead of grayscale information. Furthermore, this reduces the necessary data width of the bus for pixel data information and hence hardware complexity. Resulting differences are subsequently compared to predefined image masks that represent warning and danger areas. Figure 4 shows a top view image of a work place with possible background pattern, where possible danger and warning areas, specific to the shown work piece are marked by regions of different brightness for a better visualization. Further independence from illumination influences is obtained by additional binary morphological operations, namely dilation of the edge image and

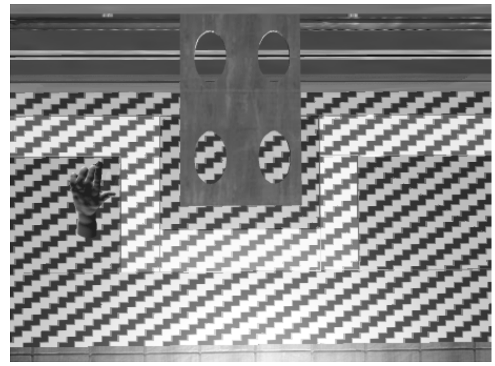

Figure 4: Synthetic top view image of a work place with possible danger areas

erosion of the difference image, each performed by a 3x3 pixel structuring element (cf. [1]).

The above-described sequence of processing steps can be implemented, following the guidelines introduced in section 3, as proposed in [2]. The Sobel operator (cf. [3]) uses a 3 x 3 pixel filter mask containing the values ± 2 , ± 1 and 0, which can be transferred directly to an FPGA implementation using usual 2D-FIR filter for serial image data as shown in figure 5 (cf. [2]). The delays T_1 and $T_2$are either implemented by using

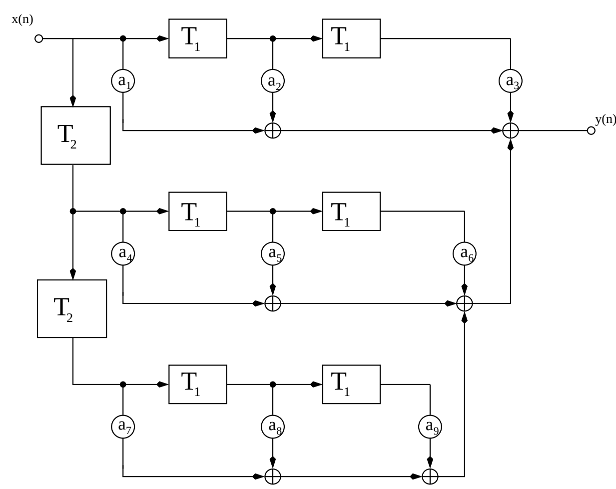

Figure 5: 2D-FIR filter for processing of serial image data

LE-latches, or external memory. Several FPGAs allow the conversion of one LE into a 16-bit shift register, which leads to the possibility of combining several LEs to achieve a long delay T_2. By exploiting these abilities, all necessary delays can be implemented using 432 LEs for images of 384 x 288 pixel and 8-bit resolution of the input signal. The Sobel operation can furthermore be simplified by replacing the squareroot expression of the gradient calculation by an absolute sum and the search for a maximum by a threshold operation. The full im-

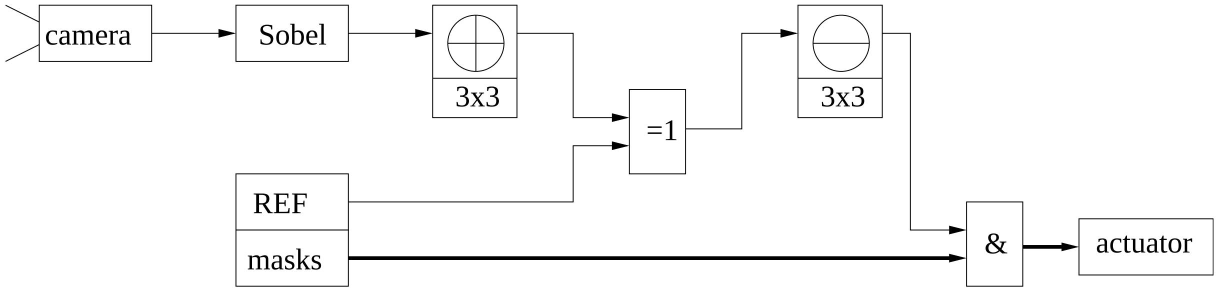

Figure 6: The processing cycle of "surveillance of a danger zone"

plementation of the simplified Sobel operator for the mentioned input signal needs only 532 LEs. Morphological operations can also be implemented in a similar 2D-FIR filter structure (cf. [4]) to act in the same way as the Sobel operator, except that signal types are binary, which drastically reduces the necessary hardware effort. Only 100 LEs are used to implement both morphological operations. Comparison with filter masks is done by a bit-wise AND concatenation of the output signal with suitably synchronized read out mask information. The whole processing cycle is shown in figure 6. The masks as well as the reference image are stored in external memory, since BlockRam is to small to hold all this information. The length of each mask word is determined by the necessary mask configuration, i.e. how many different areas have to be distinguishable. Timing simulations of the above-described synthesized implementation can be performed at frequencies of up to 86 MHz. The number of latency clock pulses is 1184, which leads for 50 Hz cameras and the above mentioned image size to a total recognition time of 20,22 ms, i.e. processing time for one image frame after image acquisition is about $220\mu s$.

5 Detection of Hands

Protection of extremities comprises the detection of hands, which are the most endangered extremities at press brakes. Accordingly, skin color detection methods can be applied for this objective. More reliable model based methods seem to require too much computational power. Skin colour detection is widely investigated for face recognition and gesture interpretation purposes. Common research areas include illumination dependency of color detection. For assessing minimal hardware effort in the present paper, application conditions are limited to a non-varying lighting spectrum and favorable background colors. The contents of the image can be reduced in a manner that disadvantageous colors are avoided. Figure 7 shows the pixels from figure 3 marked in RGB color-space (cf. [2]). It shows that skin

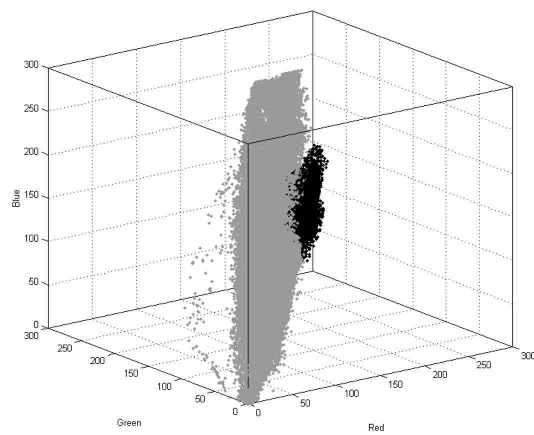

Figure 7: Pixels from figure 3 in RGB color space

color is different from other colors in the shown example. Significance can furthermore be enhanced by coercing workers to wear colored gloves. Rough distinction between skin and other colors could be applied in the presented example in RGB space by using a linear plane for separation.

A perceptron (cf. figure 8) is a very simple ANN consisting of one neuron, which establishes such kind of discriminating plane using a very simple algorithm for learning of input weights (cf. [5]). This leads to detec-

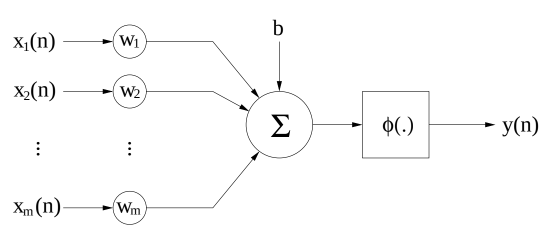

Figure 8: Perceptron

tion of skin color in a non-optimal, but "hardware efficient" way. Figure 9 shows the result of a skin color detection performed in the above presented, imperfect way.

Figure 9: Result image of the imperfect skin color detection

Several misclassified pixels have to be eliminated by considering regional information using a binary morphological operation. Figure 10 shows the respective result image after morphological erosion. The latter is compared to image masks that represent danger and warning zones as proposed in section 4.

Figure 10: Result image after morphological erosion

Further processing is necessary to consider detectability of hands and foresighted danger recognition. Both aspects can be used to enhance the safety device with respect to certainty. Evaluation of these properties however, doesn't need to take place at video frame rate, since respective attributes don't change very fast. For this reason, detectability of hands can be measured evaluating size and number of detected skin regions. Preventive danger detection can be implemented by using tracking and prediction techniques. The selection of algorithms has to be adjusted with respect to the announced kind of operation of the press.

Skin color detection by means of the presented perceptron requires at least a resolution of 11 bits for the input weights. A VHDL coded perceptron with three 8-bit input signals can be synthesized (by LeonardoSpectrum) for implementation by 444 LEs.

An adequate morphological filter requires 50 additional logic elements (cf. [4]). Comparison with appropriate filter masks is done by bit-wise AND concatenation of the output signal with suitably synchronized read out mask information. Like in the implementation of the first algorithm, masks have to be stored in external memory, using 8-bit words for each mask pixel. Detection and mask-comparison functionality can be implemented using 564 logic elements. Figure 11 shows the whole processing flow of the described algorithm. A timing simulation of the synthesized VHDL-models allows frequencies of 100 MHz using a low-cost Xilinx SpartanII FPGA device. The number of latency clock pulses is 397, which for 50 Hz cameras and the above mentioned image size of 384 x 288 pixels leads to a total recognition time of 20,08 ms. Thus, processing time is only about $80\mu s$ for one image frame.

5.1 Combination of Implementations

A combination of both implementations leads to an improvement of detection certainty. The hardware of implementation I is therefore fed by one of the three RGB camera signals instead of a monochromatic one, which would have to be evaluated first. Information from both approaches can be fused for further checking of plausibility.

6 Conclusion

Careful analysis of the workflow at press brakes delivers a-priori information which permits the implementation of a video based safety device. In particular, this strategy allows a low-cost realization of the safety device. The algorithms presented have been proven to be very efficient with respect to hardware effort. Thus, they are ideal components for implementation of a video based industrial safety application. Due to its simple structured processing paths, hardware can be designed in a cost effective manner. The robustness of hand detection can be increased by exploiting and spacial coherence consequently, modular and thus scalable versions can be designed which can be adapted to the special requirements of a particular field of application.

References

[1] William K. Pratt, *Digital Image Processing*, Wiley-Interscience, New York, USA, second edition, 1991.

[2] Jörg Velten and Anton Kummert, "FPGA-Implementation of signal processing algorithms for video based industrial safety applications,"

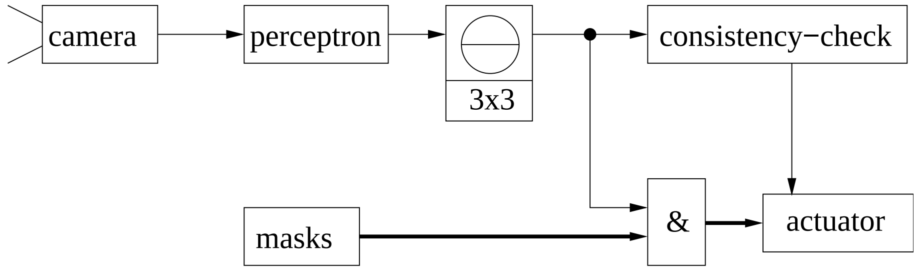

Figure 11: Processing cycle of the hand detection and region verification

in *Proceedings of the International Conference on Field Programmable Logic and Applications, (FPL'03)*, Lisbon, Portugal, 2003, pp. 1004–1007.

[3] I.E. Sobel, *Camera Models and Machine Perception*, Ph.D. thesis, Stanford University, Stanford, USA, 1970.

[4] Jörg Velten and Anton Kummert, "FPGA-Implementation of high speed n-D binary morphological operations," in *Proceedingts of the IEEE International Conference on Acoustics, Speech and Signal Processing, (ISCAS'03)*, Bangkok, Thailand, 2003.

[5] Simon Haykin, *Neural Networks*, Prentice Hall, New Jersey, USA, 1999.

[6] Werner Lehner, Anton Kummert, and Jörg Velten, "Device and method for monitoring a detection range on an operating tool," Application for US patent, 2001.

[7] F. Rosenblatt, "The perceptron: a probabilitsic model for information storage and organization in the brain," *Psycological Review*, vol. 65, pp. 386–408, 1958.

[8] John G. Ackenhusen, *Real time signal processing : design and implementation of signal processing systems*, Prentice Hall, Upper Saddle River, USA, 1999.

[9] IEEE Computer Society Press, *Proceedings of the Third IEEE Workshop on Applications of Computer Vision WACV'96*, Sarasota, USA, 1996.

[10] IEEE Computer Society Press, *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision WACV'98*, Princeton, USA, 1998.

[11] IEEE Computer Society Press, *Proceedings of the Fifth IEEE Workshop on Applications of Computer Vision WACV'00*, Palm Springs, USA, 2000.

[12] G. Matheron, *Random Sets and Integral Geometry*, Wiley, New York, USA, 1975.

[13] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, London, Great Britain, 1982.

[14] Hugo Hadwiger, *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie*, Springer Verlag, Berlin, 1957.

[15] Hermann Minkowski, "Volumen und Oberfläche," *Mathematische Annalen*, vol. 57, pp. 447–495, 1903.