

## A phase-based stereo vision system-on-a-chip

Javier Díaz<sup>a,\*</sup>, Eduardo Ros<sup>a,1</sup>, Silvio P. Sabatini<sup>b,2</sup>, Fabio Solari<sup>b,3</sup>, Sonia Mota<sup>c,4</sup>

<sup>a</sup> Department of Computer Architecture and Technology, University of Granada, Spain

<sup>b</sup> Department of Biophysical and Electronic Engineering (DIBE), University of Genoa, Via Opera Pia 11A, I-16145 Genova, Italy

<sup>c</sup> Department of Computer Science and Numerical analysis, University of Cordoba, Spain

Received 28 February 2005; received in revised form 8 July 2006; accepted 15 July 2006

### Abstract

A simple and fast technique for depth estimation based on phase measurement has been adopted for the implementation of a real-time stereo system with sub-pixel resolution on an FPGA device. The technique avoids the attendant problem of phase warping. The designed system takes full advantage of the inherent processing parallelism and segmentation capabilities of FPGA devices to achieve a computation speed of 65 megapixels/s, which can be arranged with a customized frame-grabber module to process 211 frames/s at a size of 640 × 480 pixels. The processing speed achieved is higher than conventional camera frame rates, thus allowing the system to extract multiple estimations and be used as a platform to evaluate integration schemes of a population of neurons without increasing hardware resource demands.

© 2006 Elsevier Ireland Ltd. All rights reserved.

**Keywords:** Stereopsis; Real-time; Bio-inspired systems; FPGAs; Gabor filters

### 1. Introduction

Stereo vision allows many biological systems to reconstruct depth information encoded within multiple

images. This task is accomplished in the visual cortex by a specialized receptive field structure (DeAngelis et al., 1991).

Significant studies have shown that a substantial proportion of neurons in the striate and extrastriate cortex of monkeys have stereoscopic properties; that is, they respond differentially to binocular stimuli, thus providing cues for stereoscopic depth perception (Hubel and Wiesel, 1962; Barlow et al., 1967; DeAngelis et al., 1998). Stereoscopic neurons display disparity selectivity and correlation selectivity. Many neurons have tuned disparity response profiles that collectively cover the entire range of physiological disparities. These cells can be classified on the basis of their responses: first, neurons with peak responses at (or about) zero disparity (“tuned zero neurons”, excitatory or inhibitory) which have narrow and symmetrical receptive fields; second, neurons that are tuned to larger disparities, either crossed (tuned near neurons) or uncrossed (tuned far neurons). These have broader excitatory receptive fields that are

\* Corresponding author at: Department of Computer Architecture and Technology, University of Granada, ETSI Informática, c/ Periodista Daniel Saucedo Aranda s/n, E-18071 Granada, Spain.

Tel.: +34 958 240461; fax: +34 958 248993.

E-mail addresses: [jdiaz@atc.ugr.es](mailto:jdiaz@atc.ugr.es) (J. Díaz), [eros@atc.ugr.es](mailto:eros@atc.ugr.es) (E. Ros), [silvio.sabatini@unige.it](mailto:silvio.sabatini@unige.it) (S.P. Sabatini), [fabio.solari@unige.it](mailto:fabio.solari@unige.it) (F. Solari), [smota@uco.es](mailto:smota@uco.es) (S. Mota).

<sup>1</sup> Present address: Dept. Arquitectura y Tecnología de Computadores, University of Granada, ETSI Informática, c/ Periodista Daniel Saucedo s/n, E-18071 Granada, Spain. Tel.: +34 958 246128; fax: +34 958 248993.

<sup>2</sup> Tel.: +39 010 3532289/794; fax: +39 010 3532289/777.

<sup>3</sup> Tel.: +39 010 3532289/794; fax: +39 010 3532289/777.

<sup>4</sup> Present address: Area de Ingeniería de Sistemas y Automática, Dpto. de Informática y Análisis Numérico, Universidad de Córdoba, Campus de Rabanales, E-4071 Córdoba, Spain. Tel.: +34 957 21 21 72; fax: +34 957 21 86 30.

asymmetrically wider toward the smaller disparities, and commonly include an inhibitory component around zero disparity. Other stereoscopic cells have reciprocal profiles (“near” or “far” neurons, respectively) in the sense that they respond with excitation to crossed or uncrossed disparities and with suppression to disparities of the opposite sign (Poggio et al., 1988).

Furthermore, binocular depth perception is useful in many visual applications such as autonomous robot navigation and grasping tasks. Due to the intensive calculation required to estimate the disparity values, most of the approaches implemented so far process the sequences off-line, rendering them unsuitable for real applications. The use of customized hardware allows us to process stereo-image sequences in real-time. These hardware-based approaches generally use correlation-based models (Brown et al., 2003) because they are quite suitable to hardware architecture. In contrast to feature correspondence and correlation techniques, during the last decade phase-based computational models have been proposed as an interesting alternative (Fleet and Jepson, 1993; Fleet et al., 1996), mainly because they are based on local operations and produce dense depth maps with direct sub-pixel resolution. Several real-time approaches based on this technique have recently been proposed by Porr et al. (2002) and Darabiha et al. (2003).

In this paper we describe how to deal with the properties of bio-inspired systems to be designed as embedded systems for real-world applications. We describe an embedded stereo processing system based on an FPGA device known as a system-on-a-chip (SoC), which computes a modified phase-based technique originally described by Solari et al. (2001). This model avoids the explicit computation of the single local phases of Gabor-filtered binocular images, making the approach hardware friendly and thus allowing our design to outperform previous approaches. The system includes all the hardware controllers necessary for a two-camera frame-grabber, external memory management units, VGA visualization output generation, user control interface for system configuration, etc. This allows us to use it as a smart embedded sensor that works as a system-on-a-chip, providing low level vision disparity information.

## 2. From biological models to real-time hardware systems

Engineering processing architectures designed for tasks that biological systems solve with impressive ease can benefit considerably by mimicking computing strategies developed by nature over long periods of evolution. But the adaptation of such techniques is not straight-

forward, since the physical principles upon which biological tissues are based are very different from those characteristically used in electronic technology. Furthermore, biological and electrical “technologies” face different restrictions which are overcome by resorting to different strategies.

Nevertheless, an “opportunistic attitude” which takes the key-functional principles that contribute to the outstanding performance of biological systems and also uses technology-motivated computing techniques to adapt those computing primitives must be of considerable interest. This opportunistic approach should on its own merits provide a suitable solution to the individual task in question, whilst also helping to identify and characterize the functional principles that support the high performance observed in biological systems. For example, biological systems widely use massive parallel processing to overcome the slow chemical-based principles that support most of the computing and transmission principles of neurons. On the other hand, whereas electrical technology allows faster devices (more than three orders of magnitude), the connectivity allowed by current technology is restricted to 2D patterns and so this massive parallelism becomes impossible to adopt in electronic devices.

To be able to adopt biologically inspired processing schemes we use a time-slicing technique and we have developed a very fast computing unit that abstracts the functional principles upon which the emulated scheme is based. In this way, we can process in stereo the disparity between two images several times (with different shifts and spatial scales) and thus obtain multiple disparity estimations which in a biological system would have been extracted by different populations of neurons. We then integrate all these estimations constructively to achieve the best performance.

We illustrate here one example of such an approach. We have developed a very fast disparity estimation system that is able to obtain multiple disparity estimations (up to eight) at a conventional camera frame rate and VGA resolution. This allows the exploration of integration schemes in the framework of real-time processing tasks. In Section 6 we call neural population coding the set of estimations obtained on multiple scales and with multiple shift profiles. It is documented that the performance of biological systems is based upon multiple estimations (Fleet et al., 1996) and an efficient selection mechanism that integrates complementary information from different sources.

Conventionally, parallel processing of different circuits is limited due to the limited transmission bandwidth. Especially significant are the constraints deriving

from the external memory access; which is usually one of the important bottlenecks for FPGA processing capability, but due to the on-chip system management of external and internal memory, and since the described architecture consists of one single processing unit, with the whole system implemented on the same device (as a SoC), the access control is carefully designed and this bandwidth limitation is overcome. Furthermore, the proposed scheme is scalable; since there are plenty of available computing resources on the same chip, two or more processing units can be used, if further parallelism is needed, to extract more estimations or increase the spatial resolution.

### 3. Hardware-friendly phase-based stereo

In our approach, we will use only tuned-excitatory neurons. The results of Jones' and Palmer's experiments (Jones and Palmer, 1987) suggest modelling the shapes of the RFs by two-dimensional Gabor filters with variable spatial phase. In particular, experiments carried out by Pollen and Ronner (1981) suggest that most of the simple cells can be combined in pairs, one cell of each pair with even symmetry and the other one with odd symmetry. This can be modelled by a cosine function and a sine function, corresponding to the real and imaginary parts of a complex-valued Gabor filter, respectively. Among various computational vision models that make use of Gabor functions such as localized spatial filters or as basis functions for image transformations (Daugman, 1985; Porat and Zeevi, 1988; Fogel and Sagi, 1989; Chang and Chatterjee, 1993), phase-based approaches for stereo vision have been widely studied recently (Sanger, 1988; Fleet et al., 1991). In these models disparity is computed as the one-dimensional shift along the epipolar lines necessary for aligning the phase values of the bandpass filtered versions of the binocular stereo signal. An illustrative scheme is shown in Fig. 1.

Formally, the left and right observed intensities from the two eyes,  $I^L(x)$  and  $I^R(x)$ , respectively, result related as

$$I^L(x) = I^R[x + \delta(x)] \quad (1)$$

where  $\delta(x)$  is the (horizontal) binocular disparity.

Disparity can be estimated in terms of phase differences in the spectral components of the stereo-image pair (Fleet and Jepson, 1993; Fleet et al., 1996). Since the two images are locally related by a shift, in the neighbourhood of each image point the local spectral components of  $I^L(x)$  and  $I^R(x)$  are related by a phase difference equal to  $\Delta\phi(k) = \phi^L(k) - \phi^R(k) = k\delta$ , where  $\phi$  is the image local phase at this position and  $k$  is the spatial frequency.

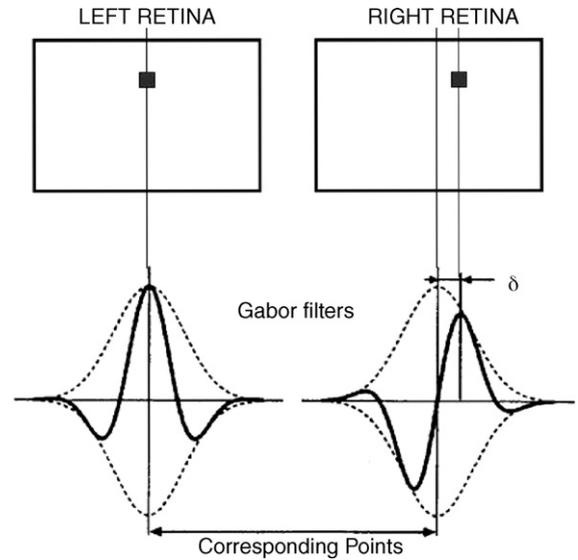


Fig. 1. Phase-based disparity estimation using neurons with receptive fields as quadrature Gabor filters.

Spatially localized phase measures can be obtained by filtering operations with complex-valued quadrature pair bandpass kernels (e.g. Gabor filters), approximating a local Fourier analysis on the retinal images (see Solari et al., 2001; Fleet et al., 1991, 1996; Fleet and Jepson, 1993). Considering a complex Gabor filter with a peak frequency  $k_0$  and a spatial extension  $\sigma$ :

$$\begin{aligned} h(x; k_0) &= \exp\left(-\frac{x^2}{\sigma^2} + jk_0x\right) \\ &= h_C(x; k_0) + jh_S(x; k_0) \end{aligned} \quad (2)$$

the resulting convolutions with the left and right binocular signals can be expressed as

$$\begin{aligned} Q(x) &= \int I(\xi)h(x - \xi; k_0) d\xi = C(x) + jS(x) \\ &= \rho(x) e^{j\phi(x)} \end{aligned} \quad (3)$$

where  $\rho(x)$  and  $\phi(x)$  denote their amplitude and phase components, and  $C(x)$  and  $S(x)$  are the responses of the quadrature filter pair. Local phase measurements are stable and with a quasi-linear behaviour over relatively large spatial extents, except around singular points where the amplitude of  $Q(x)$  vanishes and the phase becomes unreliable. This property of the phase signal yields good predictions of binocular disparity by

$$\delta(x) = \frac{[\phi^L(x) - \phi^R(x)]_{2\pi}}{k(x)} = \frac{[\Delta\phi(x)]_{2\pi}}{k(x)} \quad (4)$$

where  $[\cdot]_{2\pi}$  denotes the principal part of its argument, i.e.  $[\cdot]_{2\pi} \in (-\pi, \pi)$  and  $k(x)$  is the average instantaneous

frequency of the bandpass signal, measured using the phase derivative from the left and right filter outputs ( $x$  subscripts indicates differentiation along the  $x$ -axis):

$$k(x) = \frac{\phi_x^L(x) + \phi_x^R(x)}{2} \quad (5)$$

As a consequence of the linear phase model, the instantaneous frequency is generally constant and close to the tuning frequency of the filter ( $\phi_x \approx k_0$ ), except near singularities where abrupt frequency changes occur as a function of spatial position. Therefore, a disparity estimation at a point  $x$  is accepted only if  $|\phi_x - k_0| < k_0\tau$ , where  $\tau$  is a proper reliability threshold.

It should be noted that Eq. (4) does not require the explicit calculation of the left and right phases. Therefore, following the approach proposed by Solari et al. (2001), we can compute directly the phase difference in the complex plane using the following identities:

$$\begin{aligned} [\Delta\phi(x)]_{2\pi} &= [\arg(Q^L Q^{*R})]_{2\pi} \\ &= \arctan2(\text{Im}(Q^L Q^{*R}), \text{Re}(Q^L Q^{*R})) \\ &= \arctan2(C^R S^L - C^L S^R, C^L C^R + S^L S^R) \end{aligned} \quad (6)$$

where  $Q^*$  denotes the complex conjugate of  $Q$ .

This formulation is computationally simple because it is composed primarily of algebraic combinations of the filter outputs. Moreover, it embeds the calculation of the principal part of phase differences, without explicit manipulations of the two phases of the left and right images. In this way, it takes into account the periodicity of the phase without incurring in the “wrapping” effects on the resulting depth map. Furthermore, following (Fleet et al., 1991), for the expression of the average spatial frequency (5), to eliminate the need for an explicit calculation of phases and, consequently, the problems arising from phase unwrapping, we use the following identities:

$$\phi_x = \frac{\text{Im}[Q^* Q_x]}{\rho^2} = \frac{S_x C - S C_x}{C^2 + S^2} \quad (7)$$

where  $Q_x$ ,  $C_x$ , and  $S_x$  are the spatial derivatives of  $Q$ ,  $C$ ,  $S$ .

This approach has several advantages which make the system hardware-friendly. Although Eq. (6) increases the number of multiplications, current FPGA devices include embedded multipliers making this technology of specific interest for vision tasks. In fact, the main advantage provided by this approach is to avoid the explicit logic required for wrap-around mechanism. This implies a considerable reduction of comparison logic. Furthermore, the division operation is reduced by 50%. This

corresponds to a real benefit because the division in the fix-point arithmetic requires high precision. Although from a computational point of view there is no difference between computing disparity from differences of the phase on the monocular images or from a direct measure of the binocular phase difference (without explicit computation of monocular phases), quantization errors make the former approach noisier, which in addition requires more hardware resources. We evaluated both methods using random-dot stereograms and fix-point data of 32 bits, obtaining direct phase computation yields for higher performance when the available operation precision is limited.

To address the hardware implementation of this approach the basic steps can be summarized as follows

1. dc component image removal using the local contrast  $I - I_{\text{mean}}$  operator in a  $9 \times 9$  pixel window.
2. Even and odd Gabor 17 taps filtering of left and right images.
3. Direct phase difference calculation using Eq. (6).
4. Disparity computation using Eq. (4) assuming  $k(x) \approx k_0$ .

The dc component image removal is particularly relevant because (in a first approximation) the retina produces a “neural image” of local contrast (Shapley and Enroth-Cugell, 1984).

#### 4. Hardware system implementation

The implementation of the previous simplified phase-based model (Solari et al., 2001), requires being consistent with the discussion in Section 2. Large neural populations are not suitable for implementation in hardware because the available hardware resources are limited. We have designed a processing unit using fine-grain parallelism resources based on highly pipelined structures and short processing times. We describe the implementation of a SoC for real-time stereo computation which can be used in embedded systems. The device is a general purpose system for image stereo computation where the technology is based on re-configurable hardware (FPGA).

The choice of a phase-based stereo approach is also justified because of its robustness to illumination changes. As commented in (Cozzi et al., 1997), the contrast test shows that this approach is not very sensitive to differences in such magnitude. The approach seems to be rather robust to unbalanced images as well (usual in real cameras which have different luminance gain).

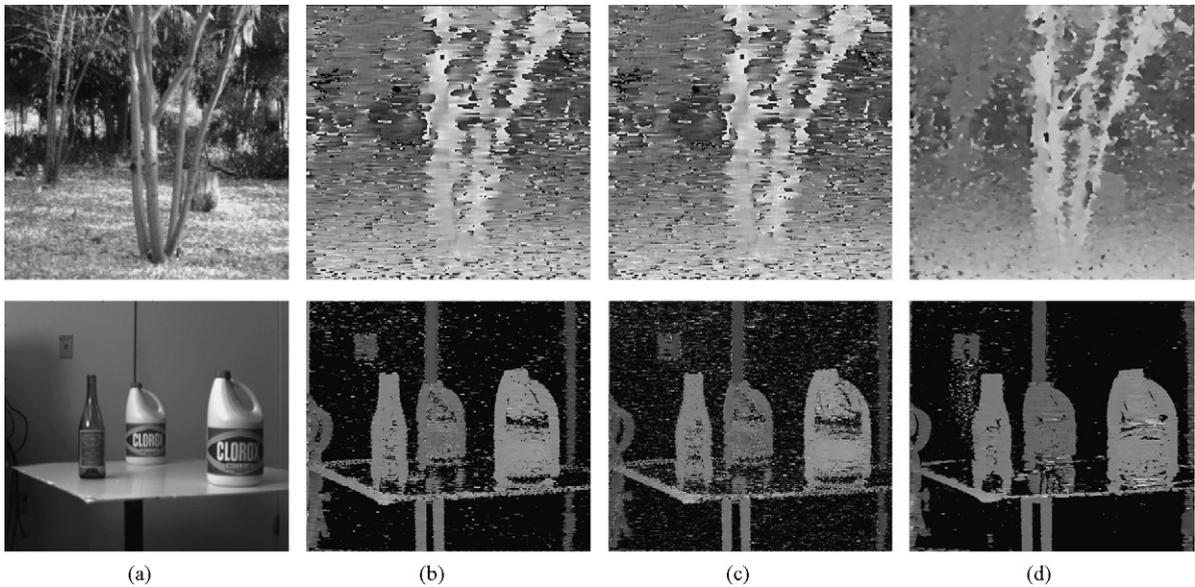


Fig. 2. Software vs. hardware implementation. (a) Original images, (b) software stereo processing, (c) hardware stereo processing, (d) results using the multiple estimation-based model described in Section 6. The disparity is encoded in grey levels, light pixels indicate short distances. Note that small differences between the software (b) and the hardware model (c) are visible as salt and pepper noise presented in the hardware produced images due to the limited precision available in the hardware implementation.

In Fig. 2 we show the algorithm outputs for a couple of standard image pairs. We compare the software and hardware results of the raw model (just one spatial scale and without neuron shifting) and we also show the results from the multiple estimation model described in Section 6.

The previous outputs (Fig. 2b and c) represent the raw data extracted from the stereo sensor encoded using a disparity-to-grey levels map. The system set-up requires image rectification and camera calibration (which is a critical stage). The present implementation only includes a simple pre-processing method based on image displacements that runs in a previous system configuration. An improved calibration pre-processing step can be implemented using an embedded calibration module to achieve better stereo-image rectification.

The hardware system architecture according to the model described in Section 1 is shown in Fig. 3.

The confidence measure used in the system is the neuron energy (module of the Gabor filter outputs) because phase is not well defined near module singularities. The system is configured by five stages in the coarse-grain pipeline (Fig. 3). All the processing stages are designed with micro-pipeline data-paths. Therefore, the total latency of the system is about 115 clock cycles. Nevertheless, the data throughput is one estimation per clock cycle. The system has been implemented in a stand-alone board as a prototype for embedded

applications, the RC300 board from Celoxica (see <http://www.celoxica.com>). All the processing operations are fully computed in the FPGA device (as a SoC).

## 5. System performance and requirements

The system frequency is 65 MHz and produces one pixel per clock cycle meaning that we can compute up to 65 megapixels/s (for instance corresponding to 211 fps of  $640 \times 480$  pixels per image, or 52 fps of  $1280 \times 960$  pixels of resolution). The system quality depends on image resolution and disparity range. The present implementation runs well for small disparities (typically values under 4 pixels for 15 taps Gabor filters). The first stage of camera calibration reduces the global image displacement and improves the local disparity range. Compared with similar recent real-time implementations Porr et al. (2002), which process at video-frame rate and Darabiha et al. (2003), which process  $256 \times 360$  pixels per image at up to 30 fps, our system outperforms these approaches.

Table 1 shows the required resources for the whole system. Note that in the convolutional stages the processing has been done with fixed point data representation of nine bits. The arctan function has been implemented using a look-up-table of 1024 address of 10 bits with 5 fractional bits and some logic to decide the sign. As shown in Fig. 2, the hardware results are similar to the software ones implemented with double data precisions

FPGA COARSE GRAIN PIPELINE

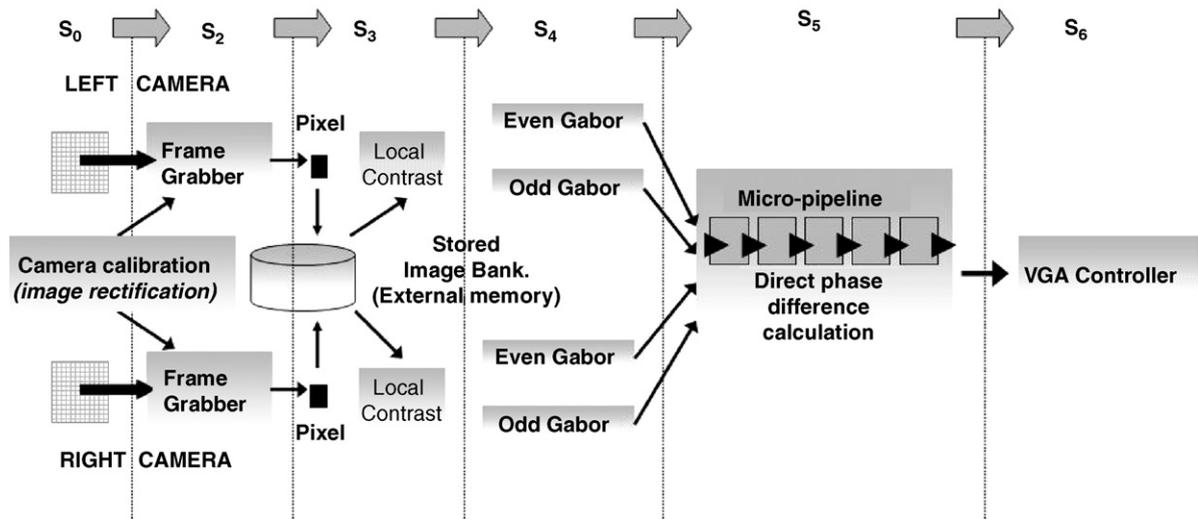


Fig. 3. Stereo hardware architecture. The figure shows the main processing units designed for the stereo vision system. Each sub-unit has been developed to process the data using a fine-grain pipeline structure. The efficient use of the intrinsic parallelism and segmentation capabilities available in the FPGAs allow the computation of one estimation per clock cycle. We have implemented a customized pipeline processing structure with parallel computing blocks in different stages for computing left and right image primitives at the same time. The micro-pipeline module computes the phase difference using a LUT for the arctan function.

and after doing several trials we consider these bit widths as good trade-offs between the system accuracy and hardware resource requirements.

Each design is characterized by the megapixels per second and is completely modular. Therefore, we can choose different resolution versus frames per second trade-off.

The FPGA re-configurability also allows different image scales computation. Provided that stereo techniques work better for small disparities, we have designed three different scales, with Gabor filters of 15, 31 and 55 taps. In this way, depending on the image structure, our FPGA can be re-configured for different scales to estimate the range of disparities that better match the image structure. Table 1 also shows the hardware resources required for these larger spatial scales

(Gabor filters of 31 and 55 taps) which enlarge the range of available disparities computable by the system but reduces their resolution. Note that the system demand grows for each scale but the computing speed in terms of fps remains constant. In future research we plan to design a multi-resolution system plus scale integration unit to compute at each pixel the scale which best fits the image properties at this position.

6. Improvements to the basic model: multiple estimation-based scheme

The main limitation of the previous system is the limited range of disparities available due to the linear approximation of the phase model. Theoretically this is  $\lambda/2$  (being  $\lambda = 2\pi/k_0$  the period of the tuning frequency

Table 1  
System resources required on a Virtex-II XC2V6000-4

Slices (%)	EMBs (%)	Embedded multipliers (%)	Mpps	Gabor spatial scale (filter taps)	Image resolution	fps
6411(18%)	15 (10%)	21 (14%)	65	15	640 × 480 1280 × 960	211 52
9197(27%)	39 (27%)	31 (21%)	65	31	640 × 480 1280 × 960	211 52
13048(38%)	71 (49%)	59 (49%)	65	55	640 × 480 1280 × 960	211 52

EMBs stands for embedded memory blocks.

Table 2  
Evaluation of the multiple estimation approach using sequences provided by Scharstein and Szeliski (2002, 2003)

	RMS error
Sawtooth	1.98
Tsukuba	1.53
Venus	1.55

of the Gabor filter) but experimentally is about  $\lambda/3$  (for details see [Cozzi et al., 1997](#)). Usually the solution found in the literature consists of a coarse-to-fine approach, using confidence values from coarse scales to warp the image at fine scales. The problem of such an approach is that wrong estimations propagate from coarse-to-fine scales. Furthermore, there is no biological evidence of such kinds of architecture in the brain ([Mallot et al., 1996](#)).

Contrary to this approach, a parallel processing of spatial scales with a fusion integration stage is more biologically plausible. In a similar way to [Fleet \(1994\)](#) the scales are processed in parallel and integrated using a similarity measure. Shift neurons could also be added ([Fleet, 1994](#); [Fleet et al., 1996](#); [Porr et al., 2002](#)) to improve the disparity range using neurons with overlapping disparity tunings. Contrary to [Fleet's](#) approach ([Fleet, 1994](#)), which uses Gabor filter correlation and sub-pixel estimations by linear interpolation, our scheme uses sum of absolute differences (SAD) over the energy of the shifted cells (which is more hardware-friendly because it avoids square roots and division operations). At this stage, the cell with the lowest response encodes the winner shift value which achieves the best disparity tuning. Phase difference for sub-pixel estimation (instead of linear interpolation methods) is used to obtain sub-pixel disparities values. The shift offset obtained with SAD, is calculated with the value obtained from the basic model providing the improved sub-pixel disparity estimation.

Qualitative results for this model are shown in [Fig. 2d](#). Note that the disparity range and resolution are improved, obtaining smooth variation and disparity details. The approach has also been evaluated numerically with the sequences used in ([Scharstein and Szeliski, 2002, 2003](#)), for which we know the ground-truth. The accuracy using the RMS (root-mean-squared) error (measured in disparity units) between the computed disparity map and the ground-truth map is summarized in [Table 2](#). The used parameters are the following: 9 shifted neurons (with a distance of 5 pixels between them) with  $\lambda = 14$  to cover a wide disparity range (from  $-24$  to  $24$  pixels) with overlapping.

The processing speed of the system using a customized frame-grabber allows us to test several population types and fusion methods in real-time. For example, we can process each image pair eight times, using three spatial scales and a shifted distribution of five neurons with overlapping disparity tuning to increase the available range of disparities obtaining an equivalent circuit running up to 26 fps of image sizes of  $640 \times 480$  pixels using approximately the same system resources (memory resources demand is increased). Shift neuron just implies offset values in the frame-grabber of one of the cameras and the different scales imply just changing the Gabor filter coefficients. Therefore, we use the same primitives described in [Section 3](#). Furthermore, the outstanding processing speed achieved by our approach allows us to use the same circuits to process the images repetitively (with different shifts and filter scales) storing the results to be integrated in a simple winner-takes-all stage. In this fusion module we just take at each pixel the disparity value (among candidates) with the highest confidence value.

## 7. Conclusions

The adopted stereo computation technique is efficient and hardware-friendly. It provides sub-pixel resolution and the disparity range can be adapted to the image structure. Furthermore, it allows, in a straightforward manner, a multi-scale and multi-shift approach as an immediate improvement. The hardware is very powerful (65 megapixels/s that can be arranged as 211 fps of  $640 \times 480$  pixels per image). This outstanding performance with a customized frame-grabber allows the system to be used as a platform for studying different models of neural population coding and integration mechanism (which take full advantage of multiple disparity estimations) in real-time tasks.

We present a way of implementing a biological model onto programmable hardware which runs on a stand-alone chip for embedded applications. The efficient exploitation of the computing resources available on FPGA devices leads to an outstanding processing speed. A customized pipeline processing structure, including some well-balanced parallel processing modules, efficiently performs phase-based stereo estimations (about one million gates on the Virtex-II FPGA are required for the 15 taps Gabor filter system). The accuracy of the system depends on the bit-width adopted at the different computing stages; this is quantified using benchmark images. Some illustrative and promising results are shown in [Fig. 2](#).

## Acknowledgments

This work has been supported by the EU grant DRIVSCO (IST-016276-2) and the National Spanish Grant DEPROVI (DPI2004-07032).

## References

- Barlow, H.B., Blakemore, C., Pettigrew, J.D., 1967. The neural mechanism of binocular depth discrimination. *J. Physiol.* 193, 327–342.
- Brown, M.Z., Burschka, D., Hager, G.D., 2003. Advances in computational stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (8), 993–1008.
- Chang, C., Chatterjee, S., 1993. Ranging through Gabor logons—a consistent, hierarchical approach. *IEEE Trans. Neural Netw.* 4, 827–843.
- Cozzi, A., Crespi, B., Valentinotti, F., Wörgötter, F., 1997. Performance of phase-based algorithms for disparity estimation. *Mach. Vision Appl.* 9 (5/6), 334–340.
- Darabiha, A., Rose, J., MacLean, W.J., 2003. Video-rate Stereo Depth Measurement on Programmable Hardware (CVPR '03), vol. 1, Madison, WI, June.
- Daugman, J.G., 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimised by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A* 2, 1160–1169.
- DeAngelis, G.C., Cumming, B.G., Newsome, W.T., 1998. Cortical area MT and the perception of stereoscopic depth. *Nature* 394, 677–680.
- DeAngelis, G.C., Ohzawa, I., Freeman, R.D., 1991. Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature* 352 (6331), 156–159.
- Fleet, D.J., 1994. Disparity from local weighted phase-correlation. *IEEE Int. Conf. Syst. Man Cybern.* 1, 48–54.
- Fleet, D.J., Jepson, A.D., 1993. Stability of phase information. *IEEE Trans. Pattern Anal. Mach. Intell.* 15, 1253–1268.
- Fleet, D.J., Jepson, A.D., Jenkin, M.R.M., 1991. Phase-based disparity measurement. *CVGIP: Image Understand.* 53 (2), 198–210.
- Fleet, D.J., Wagner, H., Heeger, D.J., 1996. Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res.* 36 (12), 1839–1857.
- Fogel, I., Sagi, D., 1989. Gabor filters as texture discriminator. *Biol. Cybern.* 61, 103–113.
- Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154.
- Jones, J.P., Palmer, L.A., 1987. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58 (6), 1233–1258.
- Mallot, H.A., Gillner, S., Arndt, P.A., 1996. Is correspondence search in human stereo vision a coarse-to-fine process? *Biol. Cybern.* 74 (2), 95–106.
- Poggio, G.F., Gonzalez, F., Krause, F., 1988. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J. Neurosci.* 8, 4531–4550.
- Pollen, D.A., Ronner, S.F., 1981. Phase relationship between adjacent simple cells in the visual cortex. *Science* 212, 1409–1411.
- Porat, M., Zeevi, Y.Y., 1988. The generalized Gabor scheme of image representation in biological and machine vision. *IEEE Trans. PAMI* 10, 452–467.
- Porr, B., Nürenberg, B., Wörgötter, F., 2002. A VLSI-compatible computer vision algorithm for stereoscopic depth analysis in real-time international. *J. Comput. Vision* 49 (1), 39–55.
- Sanger, T.D., 1988. Stereo disparity computation using gabor filters. *Biol. Cybern.* 59, 405–418.
- Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV* 47 (1–3), 7–42.
- Scharstein, D., Szeliski, R., 2003. High-accuracy stereo depth maps using structured light. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, vol. 1, Madison, WI, pp. 195–202.
- Shapley, R., Enroth-Cugell, C., 1984. Visual adaptation and retinal gain control. *Progr. Retinal Res.* 3, 263–346.
- Solari, F., Sabatini, S.P., Bisio, G.M., 2001. Fast technique for phase-based disparity estimation with no explicit calculation of phase. *Electron. Lett.* 37 (23), 1382–1383.