A Distributed Adaptive Architecture for Analog Stereo Depth Estimation

Giacomo M. Bisio[a], Bruno Crespi[c], Luigi Raffo[b]
Silvio P. Sabatini[a], Giovanni Soncini[c], Alessandro Valdes[b]
[a] Dept. of Biophysical and Electronic Engineering-University of Genoa-ITALY
[b] Dept. of Electrical and Electronic Engineering-University of Cagliari-ITALY
[c] ITC-IRST Istituto Trentino di Cultura- Pantè di Povo-Trento-ITALY

Abstract

In this paper, we present an architectural description of a visual sensor capable of extracting a dense depth map of the scene under consideration. The goal of the stereo depth perception is twofold: find an estimation of distance for each point in the image and a confidence measure of this operation. Stereoscopic depth estimation is based on disparity measurements. At functional level, in the flow of processing, we distinguish: (i) adaptive convolution with Gabor filters for measuring phase, (ii) regularization process based on local context information, (iii) a comparison between left and right phase maps to determine disparity.

The computation in step (i) occurs in two stages: the first, based on a linear space-invariant array of processing units, accomplishes a convolution filtering with kernels of even symmetry at the dominant spatial frequency of the image; the second one performs a local convolution between the image and a Gabor-like filter of given phase Such decomposition allows to map efficiently the space-variant filter over a lattice of computational nodes by combining cooperatively the outputs of the computational substrate implemented in the first stage. For step (ii), one can extract the local phase of the image at the frequency specified by the Gabor filter resorting to an adaptive phase-matching technique in space, and then regularize these results. The architecture has been validated respect to both artificial and real-world images.

1: Introduction

In the last years, progresses in neural modeling, analog circuit design, and VLSI technology have conducted to the development of embedded microsystems able to perform low-cost complex elaborations in terms of silicon area and power consumption. Due to the advantages related to the possibility of integrating on the same substrate the acquisition and elaboration stages, these systems are ideally suited to perform perceptive visual processing tasks (e.g., edge detection, motion analysis, etc. [1][2] [3]). The design process of such microsystems sometimes is guided both by neural computational models of visual perception, and by considerations on the functionality of an electronic component or on the limitations posed by VLSI implementability. Hence, in the design of a perceptual microsystem a variety of disciplines converge from computational modeling and signal processing to circuit design and VLSI microelectronics. Each of these fields poses different constraints and the solution is the result of an iterative, cross-constrained design process.

In this paper, we address the problem of defining a discrete distributed computational module for generating a dense depth-map. The method is based on the output behavior of Gabor filters of variable phase. The solution is searched imposing the constraints that the algorithm should be implemented with efficiency in analog physical structures characterized by local interconnections.

2: Phase-based stereo system

Stereo depth perception is based on a comparison between two images acquired with two cameras located at a given distance on the horizontal axis and focused on the same scene. The two images are the same only on the fixation plane (i.e., at infinity, if the axes of the cameras are parallel). If we define the disparity $d(x)$ as the displacement for which: $I_{left}(x, y) = I_{right}(x - d(x), y)$, we can state that the distance of an object in x from the fixation plane is proportional to $d(x)$. This definition does not take into account occlusions and vertical changes; however, in practice, it can be considered a good approximation for small displacement and allows us to consider the problem as a one-dimensional one: $I_{left}(x) = I_{right}(x - d(x))$ Thus for any given spatial frequency θ in x direction, one observe a phase difference $\Delta\Phi(x)$ equal to $\theta d(x)$.

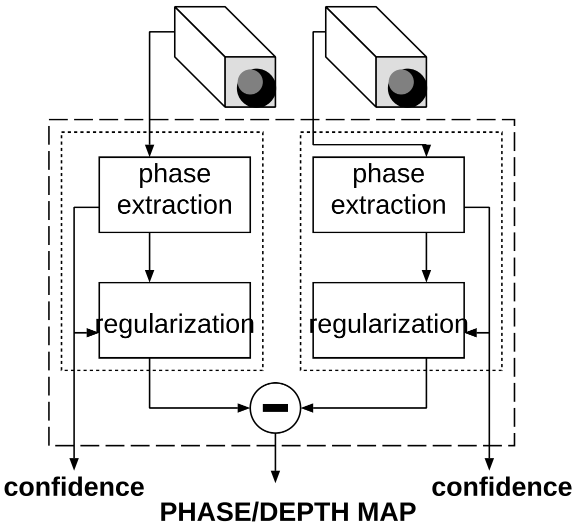


Figure 1: Block diagram of the stereo-depth system, for a one-frequency analysis. The output signals are for results and for a further decision block that merge the information at different frequencies.

The functional scheme of a stereo vision system is depicted in Fig.1. left and right eye images are processed to extract phase maps and regularize them.

At functional level, in the flow of processing, we distinguish three steps: (i) adaptive convolution with Gabor filters for measuring phase, (ii) regularization process based on local context information, (iii) a comparison between left and right phase maps to determine disparity.

3: Architectural design

We consider an architectural model discrete in space. The model is based on a space-invariant lattice network, named computational substrate. The output of this network can be locally combined to generate Gabor-like filters of different phases. The determination of the phases occurs through an adaptive approach.

3.1: Computational substrate

Many visual computations at low and intermediate levels can be thought as the result of two processes: (i) a feedforward (direct) weighted summation over local image portions performed by each output node to gain information about elementary feature elements; (ii) a cooperative interaction among neighboring nodes that propagates these local information to a larger domain for feature extraction.

In mathematical terms:

$$\sum_{k=-O} a_k y(i-k) = \sum_{k=-I} b_k x(i-k) \qquad (i \in N) \tag{1}$$

where $x(i)$ and $y(i)$ are the input and the output of the processing node at location (i), and I and O determine the neighborhood domains of the feedforward and cooperative computations, respectively. As the system is linear and space-invariant, it is fully characterized by the spatial impulse response h() (i.e., $y(i-n)$ with $x(i) = \delta(i-n)$ where (n) is a generic node). Hence the output distribution (y) is the result of the convolution between the input distribution (x) and the response to a spatial impulse (h).

When using a 1-D lattice network as specified by Eq. 1 we can approximate a Gabor function with a modulated cusp function $h(x) = Cp^{|x|}cos(\theta x + \phi)$. The relationship between the function parameters C, p, θ, ϕ and the architectural coefficients a and b can be obtained through the inverse two-side Zeta transform of the approximating Gabor function h(x). Specifically we can consider the system described by H as the cascade of two subsystems:

$$H(z) = H_1(z)H_2(z) \tag{2}$$

where

$$H_1(z) = \frac{C}{p^2z^2 - 2p(1+p^2)cos(\theta)z + 1 + 4p^2cos^2(\theta) + p^4 - 2p(1+p^2)cos(\theta)z^{-1} + p^2z^{-2}}$$

and

$$H_2(z) = p(p^2-1)cos(\theta+\phi)z + (1-p^4)cos(\phi) + p(p^2-1)cos(\theta-\phi)z^{-1}$$

$H_1(z)$ defines the recursion coefficients a_k's in Eq. 1:$a_{-2} = a_2 = p^2; a_{-1} = a_1 = -2p(1+p^2)cos(\theta); a_0 = 1 + 4p^2cos^2(\theta) + p^4$. $H_2(z)$ defines the weighted combination to obtain the Gabor-like kernel with the desired phase ϕ; this occurs when the b_k's are: $b_{-1} = p(p^2-1)cos(\theta - \phi); b_0 = (1-p^4)cos(\phi); b_1 = p(p^2-1)cos(\theta+\phi)$.

The two-stage feed-forward architectural specification, described by Eq. 2, allows to obtain tunable-phase kernels by a local (space-variant) combination (H_2) of the outputs of a space-invariant network (H_1), see Fig. 2.

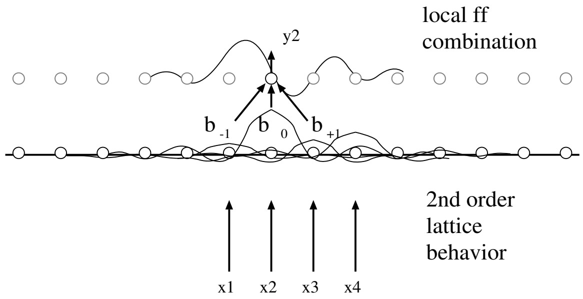


Figure 2: Schematic representation of the lattice network performing Gabor-like convolution on the input signals x

In this way, it is possible to implement, at each node of the second network, convolutions with Gabor functions with the same frequency selectivity but completely different phases (i.e. by changing the feed-forward coefficients). Moreover, it is possible to split each node of the first network into more than one node in the second network to extract different information on the same image portion.

3.2: Adaptation

According to the theory of the matched filters we want to find, for any given spatial frequency θ present in the image, the phase ϕ of the Gabor kernel for which the output $G(i;\phi)$ of the convolutor reaches its maximum. In this way, we can define a dense phase map over the lattice point as: $\Phi(i) = \{\forall i : \phi | G(i;\phi) = \max\}$. We can recast the output $G(i;\phi)$ as follows:

$$\begin{aligned} G(i;\phi) &= \sum_n I(i-n)p^{|n|}\cos(\theta n + \phi) = \\ &= G(i;0)\cos(\phi) - G(i;\pi/2)\sin(\phi) \end{aligned} \tag{3}$$

The function $G(i;\phi)$ has a maximum when $\tan\phi = G(i;\pi/2)/G(i;0)$. Actually, to avoid the direct computation of $\tan\phi$, the phase ϕ is not determined by solving the relation $\partial G(i;\phi)/\partial\phi = 0$, but through an adaptation rule based upon the steepest ascent method. The gradient of $G(i;\phi)$ function is determined and the phase is altered in the direction corresponding to the positive of the gradient. As G is invariant to a 2π shift in ϕ it is more convenient to implement the adaptive algorithm with respect to the variables $\chi \doteq \cos(\phi)$ and $\sigma \doteq \sin(\phi)$ subject to a proper constraint condition. In this way, the problem of minimizing $G(i;\phi)$ can be approached via a penalty function method. One can define the penalty function

$$P(i;\chi,\sigma) = \lambda(\chi^2+\sigma^2-1)^2 \tag{4}$$

and consider the maximization of the function:

$$F(i;\chi,\sigma) = G(i;\phi) - P(i;\chi,\sigma) \tag{5}$$

By applying the steepest ascent method to the function F one derives the following update rules for χ and σ:

$$\begin{aligned} \Delta\chi &= \eta\left[G(i;0) - \lambda\chi(\sigma^2+\chi^2-1)\right] \\ \Delta\sigma &= \eta\left[-G(i;\pi/2) - \lambda\sigma(\sigma^2+\chi^2-1)\right] \end{aligned} \tag{6}$$

where η is the computation step. It should be stressed that the iterative algorithm requires only the computation of sum and multiplication operations.

3.3: Regularization

The phase information obtained is reliable only where the output of the Gabor convolutor is strong, i.e. where the frequency component assumed for determining the phase is present in the neighborhood of the lattice point considered. In typical scenes it is likely to find critical region with poor local structure (i.e. uniform, not textured regions) while a strong signal can be detected at their boundaries. To cope with this problem, proper regularization techniques are introduced to propagate consistently the phase information from the boundary innerwise across these critical regions. The iterative procedure introduced for determining adaptively the local phase can be extended to include a regularization process over the phase map $\Phi(i)$.

We can apply regularization by imposing that the Laplacian of disparity is null over these critical region, i.e. in 1D problem, that disparity varies uniformly from one boundary to the other one. This regularization criteria implies that the phase vary with a uniform rate proportional to the spatial frequency θ:

$$\frac{d\phi(x)}{dx} = \theta \tag{7}$$

In discrete terms, the algorithm can be summarized as follows:

Function to be maximized for $i \notin \Omega_c$:

$$G(i;\chi_i,\sigma_i) = G(i;0)\chi_i - G(i;\pi/2)\sigma_i$$

Regularization soft constraints for $i \in \Omega_c$:

$$\begin{aligned}
\Phi(i) = \Phi(i+1) - \theta \quad \rightarrow \quad & \chi_i = \chi_{i+1}\cos\theta + \sigma_{i+1}\sin\theta \\
& \sigma_i = \sigma_{i+1}\cos\theta - \chi_{i+1}\sin\theta \\
\Phi(i) = \Phi(i-1) + \theta \quad \rightarrow \quad & \chi_i = \chi_{i-1}\cos\theta - \sigma_{i-1}\sin\theta \\
& \sigma_i = \sigma_{i-1}\cos\theta + \chi_{i-1}\sin\theta
\end{aligned}$$

Hard constraint:

$$\chi_i{}^2 + \sigma_i{}^2 = 1$$

4: Results

We tested the behavior of the algorithm both for artificial and natural images. We first applyed the algorithm to the synthetic stereo pair in Fig.3, in which the boxes are at different distances: the one on the right is at the same distance as the background. The resulting depth map is depicted in Fig.3: bright grey corresponds to objects closer to the viewer.

We also applyed the algorithm to the SRI-tree sequence considering the first two images as the stereo pair (see Fig.4). In this test the closest objects to the viewer are the trees on the left and their leaves on the top-right. In Fig.4 the result are presented: trees, leaves and grass are correctly detected after regularization (see Fig.4).

5: Appendix: circuital implementation of the system

The input/output relation described by Equation 1 can be thought as an infinite extent lattice network which gains its system functionality through the cooperative action of an ensemble of nodes topologically organized on a regular grid. More specifically, the overall functionality derives from the a local combination (i.e. H_2 in Eq. 2)

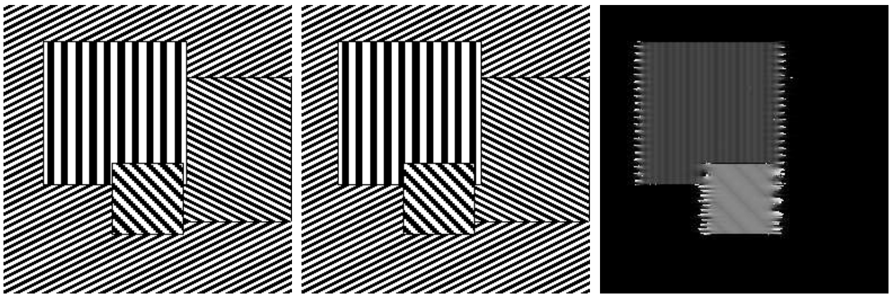
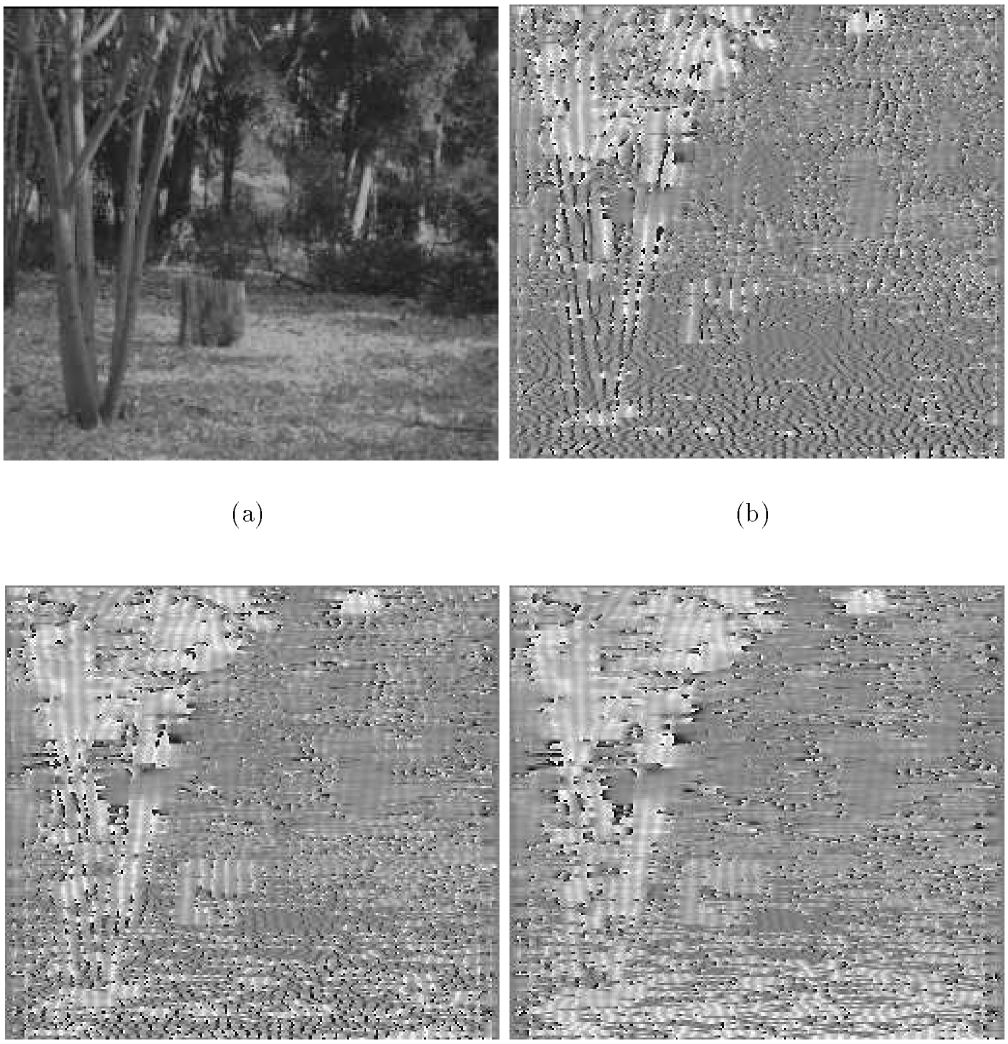


Figure 3: A pair of artificial stereo images and the result of the algorithm: bright grey corresponds to objects closer to the viewer.



(a) (b)

(c) (d)

Figure 4: (a) A SRI-tree stereo pair (only the left image is shown); (b)(c)(d) the results (depth maps) of the elaboration increasing the regularization parameter. In (d) also the depth of grass pattern is evidenced.

of the outputs of a lattice network (i.e. H_1). Given the linearity and the symmetry of the interactions between the nodes of the lattice (i.e. the a's coefficients), H_1 can be interpreted in circuital terms as a regular grid of resistors with connections among neighbors and next-neighbors. Each node is fed by a current generator with a current proportional to the input of the node [4]. The computation capabilities of these networks have been widely investigated by many researchers who pointed out the feasibility of this approach [5]. Moreover, this description can be translatet into an embedded microsystem through analog VLSI technology. In this kind of systems the current generators at each pixel site are usually photo-sensors [1][2].

Acknowledgements: This work is partially supported by the European Community under the project ESPRIT-CORMORANT 8503.

References

[1] C.A. Mead, "Neuromorphic electronic systems", *Proc. IEEE*, vol. 78, pp. 1629–1636, 1990.

[2] H. Kobayashi, J.L. White, and A.A. Abidi, "An active resistor network for Gaussian filtering of images", *IEEE J. on Solid State Circuits*, vol. 26, pp. 738–748, May 1991.

[3] C. Koch, "Seeing chips: Analog VLSI circuits for computer vision", *Neural Comp.*, vol. 1, pp. 184–200, 1989.

[4] L. Raffo, "Resistive network implementing maps of gabor functions of any phase", *Electronics Letters*, vol. 31, pp. 1913–1914, 1995.

[5] B.E. Shi and L.O Chua, "Resistive grid image filtering: Input/output analysis via the cnn framework", *IEEE Trans. on Circuit and Systems: I. Fundamental Theory and Applications*, vol. 39, pp. 531–548, 1992.