Perception of Depth Information by Means of a Wire-Actuated Haptic Interface

P. Arcara[1], L. Di Stefano[1], S. Mattoccia[1], C. Melchiorri[1], G. Vassura[2]
1. DEIS, 2. DIEM
University of Bologna - Via Risorgimento 2,
40136 Bologna, Italy
{parcara,ldistefano,smattoccia,cmelchiorri}@deis.unibo.it, gabriele.vassura@mail.ing.unibo.it

Abstract

The VIDET project is aimed at investigating the possibility of developing a wearable robotic system for helping the mobility of visually impaired persons. The basic idea consists in the conversion of real-time depth data gathered through stereo-vision into a virtual, "bas-relief" model perceivable by means of a haptic interface. In this paper we describe the real-time stereo system, review the basic principles of the two main haptic devices developed so far, and present new experimental results concerning extraction of depth data by the stereo system and haptic perception of the virtual model recovered from stereo-data.

1 Introduction

During the last years a number of interesting projects have dealt with the development of mobility aids for the visually impaired [1]-[4].

This topic is also being addressed at the University of Bologna in the context of an interdisciplinary research project called VIDET (VIdeo DEcoder by Touch). Project's objective is to investigate the possibility of developing a wearable robotic system capable of generating a virtual 3D model of the surroundings to be perceived by a visually impaired user through a tactile sensation. To this end, a complex device based on the integration of a binocular stereo-vision system and a wire-actuated "haptic-interface" has been conceived. The stereo-vision system is responsible for extracting in real-time the depth-map associated with the portion of the scene viewed by the cameras. Then, the spatially-sampled depth data obtained by the stereo system are scaled and interpolated with triangular patches in order to build a continuous "bas-relief" model of the observed scene. This model is perceived by the user through a tactile sensation provided by the wire-actuated haptic-interface, which allows to constrain the motion of her/his finger according to the shape of the model. Basically, the user should be able to interact with the surroundings by "touching" a scaled model of the scene seen by a pair of cameras.

With respect to the above referenced projects, VIDET's original approach consists in the conversion of real-time depth data gathered through a stereo system into a virtual "bas-relief" model perceivable by means of a haptic interface. So far the VIDET project has led to the development of several prototypes of the haptic-interface and of a real-time stereo vision system [5]-[8].

In this paper we describe the real-time stereo system, review the basic principles of the two main haptic devices, which have been already discussed in detail in previous work, and present new, recent experimental results concerning the extraction of depth data by stereo-vision and the perception of the 3D model recovered from stereo-data by a one-wire haptic device.

2 Depth Extraction by Stereo Vision

As mentioned in the introduction, the first step in recovering the virtual 3D model consists in extraction of the depth-map of the scene by means of a binocular stereo-vision system. This implies matching tokens between the two images captured by the acquisition system, with possible tokens being small image areas or features such as edges, corners, lines, curves. With regards to VIDET, our choice has fallen on area-based matching since, unlike feature-based matching, this approach can yield dense depth maps and hence holds the potential for embodying richer information

on scene's structure into the bas-relief model provided to the visually impaired user.

The epipolar constraint, which ensures that homologous points must lay on homologous image lines called epipolar lines, is the fundamental constraint for disambiguating the stereo-matching process . Yet, it can be exploited efficiently only when epipolar lines coincide with image scan-lines, i.e. with cameras in the so-called "standard-configuration". In the initial phase of our work this was achieved using a single camera fixed to a track and capable of horizontal translations. However, this simple solution is not suited to real-time operation since the two images cannot be acquired simultaneously. Currently, we use a Stereo-Head (STH-V1) by Videre Design [9]. This system consists of a compact motherboard with two genlocked cameras carefully aligned so as to obtain a satisfactory approximation of the standard configuration. Moreover, STH-V1 features a "line-interlace" operation mode in which alternate horizontal lines from each camera are combined into a single NTSC video signal. This allows to acquire images simultaneously from the two cameras using a single frame-grabber board.

The area-based matching algorithm developed within the VIDET project, referred to as Videt Stereo Algorithm and abbreviated as VSA, is outlined in Fig. 1.

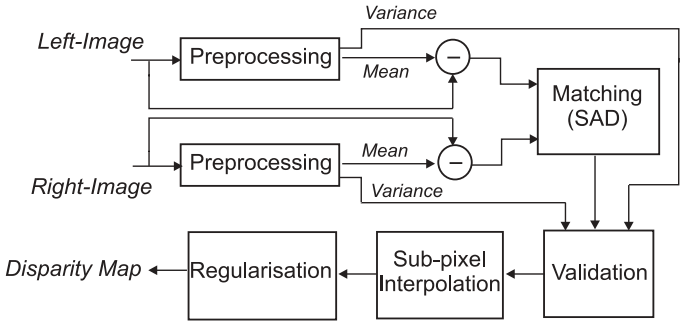

Figure 1: The Videt Stereo Algorithm.

First, the two images are preprocessed in order to compute the mean and variance of the gray-level distribution in a small window centered at each pixel. Then, the mean-value images are subtracted to the original images. This normalization step is useful to compensate for possible gray-level variations due to slightly different settings of the cameras [10].

The normalized images are then matched based on the SAD (Sum of Absolute Differences) similarity criterion. It is well known that the set of matches obtained at this stage by an area-based algorithm may be highly unreliable and therefore a validation step must be incorporated into the algorithm to increase robustness [10]- [11]. Among the major sources of matching errors are occlusions, low-textured regions and repetitive patterns.

Validation is typically done by exploiting, besides the epipolar constraint, additional matching constraints. Those included in VSA's validation step consist in checking matches' left-right consistency, in discarding matches found in poorly-textured areas and in allowing a unique match per each pixel. The left-right consistency check [12] performs very well in eliminating wrong matches due to occlusions and may partially deal with those generated in poorly-textured regions. However, as for uniform regions, the safest solution is to reject all matches found at points exhibiting poor local texture. Therefore, we estimate local texture by the variance of the intensity distribution, which is computed and stored in the preprocessing step, and discard matches found at points where the variance is below a given threshold. The stereo imaging geometry ensures that, with the exception of occluded objects, there must be a unique match for each pixel. Hence, when a correspondence between two pixel has been established we make them no longer available for further matching. This simple constraint is effective in presence of repetitive patterns.

VSA's resolution in disparity (i.e. depth) measurements was initially 1-pixel. However, several experiments aimed at evaluating the degree of tactile perception achievable through VIDET have pointed out the need of sub-pixel resolution. Therefore, an interpolation step providing disparity measurements at $\frac{1}{4}$-pixel follows validation in current VSA'a version.

The disparity maps obtained by the algorithm may still be noisy and, due to the validation process, typically contain unmatched points. Hence, the final VSA's step, referred to as Regularisation, consist in first cleaning-up noise by a median filter and then propagating the available disparity information over unmatched points by a linear scheme incorporating a disparity-gradient constraint aimed at preserving depth discontinuities.

3 Prototypes of the Haptic Device

Two main versions of the haptic interface have been developed. Common feature of these devices is that they are based on actuated wires instead of a more traditional "rigid" mechanical chain (see [13] for a survey

on similar haptic devices). This choice permits to have a light and wearable device, with a reduced power consumption.

The two prototypes differ in the number, respectively one and three, of actuated wires. Since the minimum number of wires needed to generate forces in any direction in a 3D space is 4, devices with 1 and 3 actuated wires are intrinsically defective, i.e. with less actuated degrees of freedom than those necessary, and the interaction perception is somehow limited.

Figure 2: The 3-wire interface developed for VIDET.

The three-wires device is shown in Fig. 2. Each wire is connected to a different motor, capable of applying the desired tension on it, and equipped with a suitable sensor to measure the tension. The length of each wire is computed using the encoder of the associated motor. The three wires are connected to a thimble, by which the user can touch the virtual surface. The 3D position of the thimble is determined by the wire lengths, and so forward and inverse kinematic can be easily computed [6].

Figure 3: The 1-wire interface developed for VIDET.

The one-wire interface is shown in Fig. 3. The determination of thimble's position in the 3D space is, in this case, more complicated, since the wire length is no longer sufficient. Hence, additional force sensors have been integrated into the device to compute thimble's position (see [8] for details) and such a computation is performed mainly on the basis of the measurement of the wire tension.

With both devices, a relevant issue is the computation of a proper force that must be generated to the thimble by the haptic interface for accurate perception of the virtual environment.

Ideally, the force applied to the thimble should be 0 in the case of movement in the free space (i.e. no contact with virtual objects) and should be computed according to its position and to the shape of the contact surface (ideal force) in case of interaction with a virtual object. Indeed, in case of motion in the free space a very low non-zero tension has to be applied on each wire in order to maintain a positive tension and, as for the contact with a virtual object, the ideal force could not be applicable. In fact, as already mentioned, the device is defective and one has to implement a force as similar as possible to the ideal one, see e.g. [6]. Several methods of computing this force have been considered and evaluated. The key point is the following: the one-wire device is able to generate only forces directed along the wire direction, while the three-wire device can generate forces internal to the three wires (as a sum of the wire tensions).

The haptic interfaces developed so far are based on a tension control of the wires. A suitable tension sensor is present in the actuation box, as well as an encoder which allows to know the length l of the wires. Starting from the force to be applied to the thimble one can compute the tension on the wire(s), this tension is the set-point signal for the regulator; the system input, i.e. the motor torque, is computed via a simple PID controller in order to give the desired tension on the wire(s).

Perception with the 3-wires device is better since a wider range of forces is applicable to the user; on the other hand the 1-wire device is simpler in terms of structure and control (since one has to control only a motor torque) and therefore constitutes a valid and attractive alternative. The most significant problem to be solved with the one-wire device is the exact computation of the position of the thimble in the 3D space. To this end, a new sensor is currently under development [14].

4 Experimental results

An example of the typical results obtained using the Video Stereo Algorithm is now reported.

Figure 4: Disparity map.

Fig. 4 shows the disparity map obtained by VSA on the image in Fig. 5, which is the left image of a stereo pair.

Figure 5: Left image of the stereo pair.

Disparity values are encoded with 255 gray-levels, with brighter gray-levels representing points closer to the camera. The extracted 3D information is displayed in a different manner in Fig. 6, which represents a lateral view of the VRML model built by projecting the image in Fig. 5 onto the disparity map in Fig. 4.

The algorithm is capable of recovering correctly

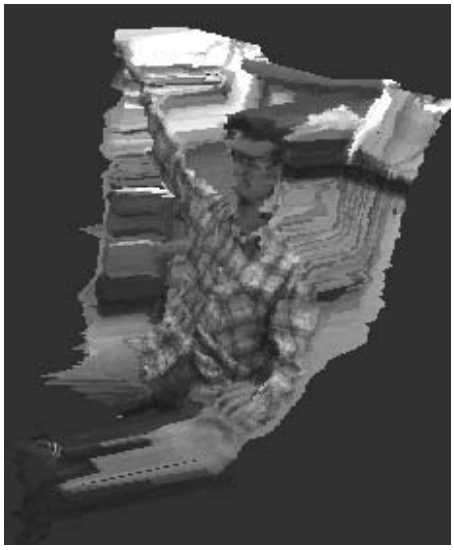

Figure 6: VRML model: rotated view.

scene's basic 3D structure: the person's figure appears closer to the camera than the background and the sharp depth gradient between the foot and the body has been detected neatly. This can be seen easily from the disparity map as well as from the VRML view. The extracted map encodes also the smooth depth variation along the person's leg and arm: this is less clearly visible in Fig. 4, where however the gray-tones gets darker along the foot to leg path, whilst is more evident in Fig. 6. The VRML view shows also the sharp depth variation between the person and the furniture behind him, while the position of this object with respect to background is perhaps easier to appreciate in Fig. 4, where the top-right area of the image is brighter than background (as well as darker than the person's figure).

VSA has been tested extensively with stereo pairs captured with the single-camera method and the Stereo-Head, as well as with standard test-images available on the Web. The results show that the algorithm is effective in generating reliable and dense disparity maps. A collection of experimental results (including original stereo-pairs, disparity maps, VRML reconstructions and animated sequences of real-time disparity estimation) can be found at the web-site http://labvisione.deis.unibo.it/~ smattoccia/.

Perception of virtual objects has already been shown and discussed in previous work (see e.g. [7, 8]) in the case of simple artificial test objects, i.e. surfaces not generated by the stereo vision system but created synthetically to test the devices. Here, we present new

experimental results concerning the exploration by the one-wire device of the 3D surface attained from depth data extracted from a real scene by the Videt Stereo Algorithm.

We consider the scene of Fig. 5 and use the bas-relief model built on the basis of the disparity map shown in Fig. 4. The experiments consist in exploring the model with the thimble and are aimed at evaluating the capability of the device to enable perception of the model's 3D shape, i.e. to properly control the thimble in order to follow accurately the reliefs of the model. To this purpose we have carried out several explorations and recorded the trace associated with the 3D position of the thimble's tip. This allows to evaluate the performance of the device by comparing the recorded trace with the known model's shape.

An example of the typical results obtained in these experiments is shown in Fig. 7, in which the trace of the thimble has been superimposed to the bas-relief model associated with a small portion of the scene of Fig. 5. The model in Fig. 7 comprises the head and the top part of the trunk, with the right and left parts of the figure flipped with respect to the view in Fig. 5. From Fig. 7 it is possible to see that the device enables to follow quite accurately the shape of the surface; in particular, one can see clearly from the trace the perception of the relief associated with the trunk.

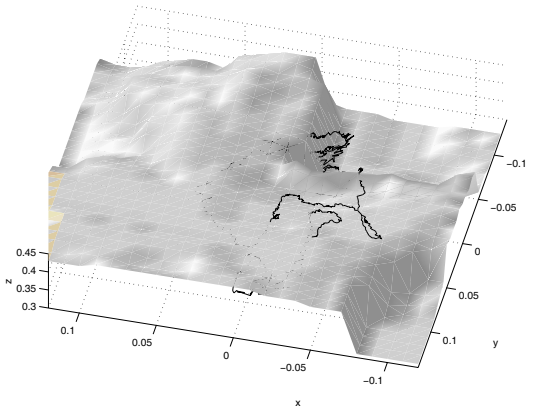

Figure 7: Exploration of a reconstructed surface with the 1-wire interface.

It is worth underlying that, as already mentioned, perception is meant here as the ability to follow accurately a surface's shape, not that of recognizing the shape. So far, experiments aimed at recognising shapes by trained users have been successfully carried out only in the case of simple artificial test-objects.

Since Area-Based stereo algorithms are computationally very expensive, one major goal of our work has been the development of optimization strategies that may render the matching process suitable to real-time processing. One particularly effective optimization incorporated into the algorithm consists in the massive use of the box-filtering technique [10], [15], which allowed us to minimise the number of operations per pixels executed in all of VSA'a computationally intensive steps. Having minimised the computation per pixel, additional speedup has been achieved by MMX-coding large parts of the algorithm in order to process several pixels in parallel. As a result, VSA can provide video-rate disparity data for the VIDET system running on a Pentium II PC.

The perception experiments reported in this section have been carried out with the stereo-vision and the haptic device controlled by two separate computers and not linked together for real-time operation yet. They may be referred to as "static" experiments, since the depth map extracted by the vision system is stored on disk to be explored at a later time with the haptic device. Very recently, thanks to the development of the highly optimised version of the stereo algorithm, the two systems have been linked together for real-time operation, i.e. VSA runs in real-time on the image stream generated by the Stereo-Head and continuously updates the depth map explored by the user with the one-wire device. The initial real-time experiments show that, besides enabling perception of the 3D structure of simple static scenes, the system allows to promptly track the variation of depth associated with motion in simple dynamic scenes. For example, if an object under exploration is suddenly moved farer from the cameras the user feels promptly that he has some free-space in front of the thimble and can move it forward until he touches the object again in the new position. Similarly, with the same suddenness, if the object is brought nearer to the cameras the user feels an increasing tension on the wire that tends to move the thimble backwards, until he can touch the object in the new position.

5 Conclusion and Future Work

We have described VIDET's main components, the stereo system and the haptic device, and presented some new results regarding the integrated functioning of these components. In our previous work, haptic perception had been demonstrated only in the case of syntectic test objects.

The new results show that the stereo system captures correctly the basic 3D structure of the scene and that this can be perceived quite accurately using the one-wire haptic device. Thanks to the development of a highly optimised implementation, the stereo algorithm is now capable of generating depth-data at video-rate. This allows for integrated, real-time operation of the overall system. Preliminary experiments with simple dynamic scenes show that the system enables prompt following of the depth variations associate with motion.

Our current work is aimed at assessing the degree of scene understanding achievable as a result of the haptic perception provided by the system. To this purpose we have started a program of experiments, currently under development, involving both normally sighted and visually impaired users. This program addresses issues such as the ability of clearly detecting and localizing objects from background, to distinguish between multiple objects, to understand object's shapes, to estimate the distances between objects and to assess basic facts in dynamic scenes (e.g..the presence of a new obstacle). The results of this activity will be presented in a future paper.

Acknowledgments VIDET is an interdisciplinary research project at the Univ. of Bologna coordinated by Prof. C. Bonivento, and developed by DEIS (Dip. di Elettronica, Informatica e Sistemistica), DIEM (Dip. di Ingegneria Meccanica) and DM (Dip. di Matematica). Financial support is provided by the Univ. of Bologna.

References

[1] J. Borenstein, I. Ulrich, "The Guide-Cane, a Computerized Travel Aid for the Active Guidance of Blind Pedrestians", *IEEE Int. Conf. on Rob. and Autom.*, ICRA'97, Albuquerque, NM, April 1997.

[2] G. Lacey, K.M. Dawson-Howe, "The Application of Robotics to a Mobility Aid for the Elderly Blind", *J. of Robotics and Autonomous Systems*, Vol. 23, No.4, 245-252, 1998.

[3] Molton N., Se S., Brady J.M., Lee D., and Probert P., "A Stereo Vision-Based Aid for the Visually-Impaired", *Special Edition of the Image and Vision Computing Journal*, 1997.

[4] P. Aigner, B. McCarragher, "Shared Control Framework Applied to a Robotic Aid for the Blind", *Proc. 1998 IEEE Int. Conf. on Rob. and Autom.*, Leuven, B, May 1998.

[5] C. Bonivento, A. Eusebi, C. Melchiorri, M. Montanari, G. Vassura, "WireMan: A Portable Wire Manipulator for Touch-Rendering of Bas-Relief Virtual Surfaces", *8th Int. Conf. on Advanced Robotics*, ICAR97, Monterey, CA, July 7-9, 1997.

[6] C. Melchiorri, M. Montanari, G. Vassura, "Control Strategies for a Defective, Wire-Based, Haptic Interface", *Int. Conf. on Int. Rob. and Syst.* (IROS'97) Grenoble, F, Sept. 8-12, 1997.

[7] C. Melchiorri, G. Vassura, P. Arcara, "What Kind of Haptic Perception Can We Get With a One-Wire Interface?", *1999 IEEE Int. Conf. on Rob. and Aut.*, ICRA'99, Detroit, MI, May 10-15, 1999.

[8] P. Arcara, C. Melchiorri, "3D Position Measurement Based on Force Sensors for a One-Wire Haptic Interface", *16th IEEE Instr. and Meas. Techn. Conf.*, IMTC'99, Venice, I, May 24-26, 1999.

[9] VIDERE DESIGN, "STH-V1 Stereo Head User's Manual", *http://www.dnai.com/ mclaughl/*, 1998.

[10] O. Faugeras et al., "Real-time correlation-based stereo: algorithm, implementation and applications", *INRIA Technical Report n. 2013*, 1993.

[11] L. Robert, M. Buffa, M. Hebert, "Weakly-calibrated stereo perception for rover navigation", *Proc. 5th. Int. Conf. on Computer Vision*, 1995.

[12] P. Fua,"Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserve Depth Discontinuities", *Proc. 12th. Int. Joint Conf. on Artif. Intell.*, 1991.

[13] R.L. Williams II, "Cable-Suspended Haptic Interface", *Int. Journal of Virtual Reality*, vol. 3, n. 3, pp. 13-21, 1998.

[14] C. Melchiorri, G. Vassura, "Development and Application of Wire-Actuated Haptic Interfaces", *Int. Journal of Intelligent and Robotic Systems*, Special Issue on "Humanoid Robotics and Biorobotics", J. Lenarcic Ed. (to be published).

[15] M. J. Mc Donnell, "Box-Filtering Techniques", *Computer Graphics and Image Processing*, Vol. 17, pp. 65-70, 1981.