2006－DBS－138（26）
2006－GN－58（26）
2006／1／27

On the Effectiveness of Employing 3D Video for the Remote Assessment of Fruits Quality

Rafael Sierra[1], Fumitaka Uchio[2], Nobukazu Iguchi[3] and Hirokazu Taki[2]
[1]Graduate School of Systems Engineering, Wakayama University
[2]Faculty of Systems Engineering, Wakayama University
[3]School of Science and Engineering, Kinki University

Abstract. In this paper, we introduce a real time stereo video application to be used as a remote tool for the assessment of quality in fruits. In previous research we have developed a real time video system in order to perform remote consultation; however, as the system only used a single video image, when applied to the evaluation of fruits quality, accurate diagnosis could not be done because the details of the surface of the observed object were not properly seen. Therefore in this work, we have developed a new real time 3D video transmission system and evaluated its usefulness in the assessment of fruits quality. We transmitted stereo video of mandarin oranges and with the help of experts on fruit growing we evaluated the system obtaining positive results

3D 映像を用いた遠隔果樹品質評価の有効性

シエラ・ラファエル [1], 内尾文隆 [2], 井口信和 [3], 瀧寛和 [2]
[1] 和歌山大学大学院システム工学研究科
[2] 和歌山大学システム工学部
[3] 近畿大学理工学部

本論文では遠隔地からの果実品質の評価に用いるツールとして 3 次元リアルタイムビデオアプリケーションを提案する。筆者らはこれまで、遠隔相談を行うためリアルタイムビデオシステムを開発してきた。しかし、このシステムは 2 次元動画像を用いるために果実の品質評価に用いた場合、観察する対象の詳細な表面の形状を表現できないため、正確な品質評価ができないという問題点があった。そこで本稿では、リアルタイム３Ｄビデオ伝送システムを開発し、果実品質評価に用いた場合の実用性について評価を行った。みかんの３Ｄ動画像を伝送し、果実栽培の専門家による品質評価実験を行ったところ、良好な結果を得た。

INTRODUCTION

Technology for stereo visualization or stereoscopy, also known as 3D vision or simply 3D, is well established in a number of industries, for instance, the industries of entertainment and computer aided design, and employed in the creation of things like virtual worlds and three-dimensional cinema. However, despite its availability since many years ago, the access to 3D technologies by the general public has been somehow restricted due to a series of factors such as the difficulties in producing 3D content, where such production was usually done using complex devices, such as high definition cameras with high precision requirements, or employing high performance machines, which were the only ones able to process the complex virtual models created using computer graphics.

The use of 3D technology has been also limited because of the capabilities of visualization of 3D products, requiring in all cases wearing especial devices, rather cumbersome as well as expensive, facts disliked by many users which would prefer a more natural interface with 3D technologies.

These requirements, nevertheless, are changing, and recently advances in technology have made possible that conventional desktop equipment become able to support 3D. Also, the introduction of devices that allow free viewing, i.e., no need for wearing especial accessories, has made possible take these technologies closer to final users. Furthermore, the increasing availability of high definition digital capture devices such as mega pixel cameras and now common place DV video cameras, has made less complicated the creation of contents.

On the other hand, the spread of the Internet and the increasing availability of high bandwidth and high speed network connections, enables new uses such as the transmission of high definition DV video [4] in applications like telemedicine, although only in controlled environments and with stringent network requirements such as delay and bandwidth.

One particular application we propose, similar to telemedicine, is the use of video for remote observation in agriculture, in order to perform consultation and diagnosis of problems in crops. Although the quality of image must be guaranteed, in our case the network conditions do not need to be, thus making the application suitable to be used in the public Internet.

Applications that used to have high requirements can be now implemented using common and currently available equipment without the need for investing in highly specialized devices. This situation made us consider providing a remote high quality video system to perform technical consultation for the development of agricultural products, and provide this system with capabilities of stereo video in order to increment the realism of transmitted images with the aim of achieving effective diagnosis.

Scope and purpose of this research

Within this research we want to find out whether is possible or not the use of stereoscopic video in providing effective remote technological assistance for agricultural development. Furthermore, we also want to find out if it is possible to do so avoiding the use of complex interfaces and taking advantage of technologies within the reach of everybody.

In past research at our laboratory, we have developed a system designed to perform real time video consultation to experts over the Internet [1], and it is used to perform inspection and observation tasks, (Fig.1). The low availability of experts for consultation and the difficulties involved in take the expert to the field, make attractive the option of real time video communication.

In our system, the remote observer relies on what the sender shows him; this is usually a single video image, which generally does not reproduce accurately the "real" characteristics of the object being observed. The main factor for this limitation is the deficiency of depth perception. Depth perception is based on many factors like perspective, shading, and fundamentally, stereoscopic vision. Human vision has a left-right perspective difference that needs to be replicated.

Conventional video only works with a single view and can not reproduce the effect of stereoscopic vision. If we add another camera we can simulate the way human eyes capture images, and if we combine the image produced with adequate display capabilities we can show a complete three-dimensional image that will improve the perception of depth, distance as well as the realism of the image of the object being observed. In some applications like observation of living things such as vegetables, the quality and realism of the video image are of preponderant importance; this is particularly true for the observation of a fruit, were its crispness and freshness are difficult to notice if not seen either directly or through a stereo system. For these reasons we believe there are advantages in the use of stereo video for remote consultation.

This system will provide the remote consultant with a three-dimensional - stereoscopic view of the object of consultation and its surroundings. The increased realism offers great advantages in the observation and diagnosis of problems.

Related Work

There are a number of solutions in the market for standard video transmission, mainly focused on the application of teleconference. There is even a commercial application [2] that used to support stereo video for teleconference although the unavailability of sources did not allow for experimentation. This was yet another reason for us to proceed with the implementation.

DESCRIPTION OF THE SYSTEM

As previous implementations, our system consists of three components: a sender application, a receiver application, and the network mechanism that connects them (Fig. 1). In this paper, we concentrate on the Sender and Receiver applications; we describe them in this section.

There are mainly two tasks to be performed; one is the capture and encapsulation of the video source images; the other one is the de-encapsulation and stereo display of those images. In doing these tasks we want to focus on avoiding complexity and on minimizing the requirements by avoiding traditionally used components and taking advantage of recent hardware, using low cost, off the shelf components, affordable to end users.

Stereo Image Acquisition

To acquire the images two video cameras are used. We place them horizontally one beside the other in a parallel arrangement, that is, with convergence on the infinite (Fig. 2). Convergence of the left and right video images for closer objects can be achieved at the receiver application by shifting the left image towards its left, and the right image towards its right, according to the user preferences (Fig. 5, 6). This configuration is believed to be the least uncomfortable for camera placement [3].

Stereo Image acquisition is done at the sender: For this we are able to use any pair of USB web cameras or preferably a pair of DV cameras (IEEE1394). We tested the system with a pair of Logicool Qcam Orbit QVR-1 web cams, and a pair of Sony HandyCam DCR-HC30 DV cameras. The Sony cameras also allow the use of USB streaming capabilities.

Each camera is used independently so, if desired, we can transmit only the output from one camera to, for example, a receiver lacking stereo display capabilities, without compromising the basic functionality of the system.

For maximum usability, we provide independence of video encoding and video compression. Therefore, the digital output from the cameras can be encoded into any video format available in the host computer. This process is done using the Windows Media subsystem for video (DirectShow, part of DirectX) through the creation of a "filter" which compresses/encodes the video streams. Using another filter, packets generated are encapsulated using the Real Time Protocol - RTP and send through the network. These filters created are connected up like "building blocks" into what is called a "filter graph" (Microsoft, 2005).

Filter graphs are a standard element for creating multimedia applications within the windows environment. For example, there are filters that read data from a file, filters which compress/decompress video streams, and filters that render video to the screen. The main advantage is that filters can negotiate the formats of data to exchange, automatically. We provide the filters we have created, so developers can build a stereo application tailored to their own needs.

Stereo Image Display

Image display is done at the receiver: The two video streams are received and reproduced at the receiver. To do so, another "filter graph" is created to, first, de-encapsulate received packets, and decode-decompress the video images. The filter graph at the receiver is connected directly to the filter graph at the sender to make this process transparent. Doing so, as explained in the previous subsection, the Windows media subsystem takes care of identifying the type of encoding-compression used at the sender and finds and applies the most suitable filters to render the images at the receiver automatically.

One extra step is needed at the receiver to show the images with a 3D effect. For this we use the Sharp LL-151 3D Liquid Crystal Display Monitor [7], Liquid Crystal Shutter- LCS Glasses may also be used). The Sharp 3D LCD Monitor, offers great independence to the viewer, i.e., to get the 3D effect it is only necessary to be in front of the screen at a regular distance. It does not require the use any especial hardware (but a newer graphic board) however it requires complex transformations to be performed on the images before displaying.

The transformations needed, require a modification in the color components of the image. As this has to be done pixel per pixel, processing two video images of 800x600 with 60 frames per second (30 each), may take near 60 million complex calculations per second to the processor in order to display properly the video on real time. This is an issue as it may well consume almost all the processing power of most recent computer hardware.

For this reason we decided to trick the graphic card into processing flat 2D images as 3D objects (textures), using a technique called pixel and vertex shaders (Nvidia, 2005). The powerful vector math and parallelized architecture make current GPUs much faster than general-purpose Pentium-like CPUs, for certain transformations where the current state at a certain point is only dependent upon the previous state for a small area around it, just like the 3D transformations we need to perform. Here we are exploiting this unintended use of the GPU, liberating the CPU for it to take care of other tasks of the operating system. We got a performance gain of 350% at full frame rate (Table 1).

Evaluation

Results and comments are obtained from experts using the system. The evaluation is then done in two parts: First, evaluation of the interface of the applications, i.e., how well can experts manipulate the devices used for capture and visualization of video. Second, evaluation of the utility and effectiveness of the stereo visualization system, i.e. how helpful is the system for providing remote assistance through the use of stereo video for assessing the characteristics of a particular agricultural product.

For our experiments we counted with the assistance of three experts with previous experience in the remote observation of crops trough the use of still images captured and transmitted using cellular telephones, accompanied with oral explanations. It is important to mention that the experts had also previous experience in the use of three-dimensional display systems (Liquid Crystal Shutter glasses).

RESULTS AND DISCUSSION

The product used for observation was Mandarin Orange. To assess the quality of mandarins when directly observed, according to the experts, the most important clue is provided by the skin characteristics, particularly by how accentuated is the unevenness of the surface; they look for bumps and valleys, crispness and freshness.

Inside the laboratory we set up an observation area to place the products to be evaluated, arranging the cameras as described in the previous section, placed the product at regular eye distance from the camera objectives and provided adequate lightning for correct observation (Fig. 3). Network conditions are not being evaluated this time so they are set to ideal. The experts successively used the system to observe fruits with different characteristics.

Comments on the usage of the application and discussion

The following comments where provided by the experts during and after a session of usage:

- It is a good thing being able to observe the objects in the screen with the naked eye, i.e., avoiding wearing glasses and being able to see directly. (Fig. 4)
- It is a good thing that with the calibration function, the focus of the area to be observed can be changed and not remaining with only one objective point. Calibration is good but it would be even better if were done automatically. (Fig 5.6)
- Running on Windows assures, up to a certain level, easiness of use and wide availability.
- It is little hard to adjust the position of one's head in order to observe properly the objects displayed in the 3D display. (Fig. 4)
- Continuously watching the 3D display generates visual discomfort and eye fatigue.

Observations:

It is true that when first used the screen, is a little hard to find the right position in front of it. However, from our experience working with it, we can assure that with little training one can learn to rapidly place oneself on the right position. Moreover, the monitor provides on screen clues to indicate whether we are seeing properly.

Tiredness is felt due to the eyes fixed focus on the screen surface. After a short time of using the screen this tiredness can be overcome. 3D mode is normally used in precise and short moments; the 3D screen supports switching between Stereo Mode and Normal Mode with the switch of a button.

Comments on the utility of the stereo video and discussion

The following comments where provided by the experts:

- Using Stereo Vision is appropriate as an element of judgment for the evaluation of fruits, and maybe some other products.
- The unevenness - protuberances and hollows - of the surface could be clearly seen.
- That unevenness in the surface seems to be more notorious in the screen than when holding the product in one hand, and observing it directly. The condition of the surface can be clearly seen, this is, its general aspect such as crispness and freshness can be judged with confidence.
- The scars surface can be clearly seen. Scares from early development can be easily differentiated from scares consequence of harvesting.
- The realism of the image is high, so it seems that may be even possible to observe the fuzzy down of peach skin.
- From the observation of mandarins can be concluded that it should possible the observation of strawberries and see them more clearly.
- Until now, image 2d and oral explanations were used, however, if using this three-dimensional image system, a better and concrete understanding can be achieved
- In the case of making practical use of the system, a real scenario doesn't come to mind, taking the capture equipment to a crop field is yet to be seen.
- White discoloration on leaves was already observed with enough detail with two-dimensional system, so there is no real advantage in observing those parts.

Observations:

According to this comments, the effectiveness of the 3D video is clear. It remains pending though, taking the application from the controlled environment of the laboratory to the in field experimentation.

CONCLUSIONS

Low availability of experts for consultation in the agricultural field and the difficulties involved in moving them to the field, make attractive the option of using high quality stereo video communication for assessing development of fruits and vegetables.

It is possible to build a system that allows video transmission and visualization in 3D using low cost, off the shelf components, readily available in the market.

Our system can be considered effective. 3D is effective to do evaluation of fruits using video and a network connection. The system facilitates and enhances experts labor by allowing them to work remotely saving time and other resources.

We have developed our stereo video system and concentrated on the sender and receiver applications. At the Sender, we packetize two video streams with independence of encoding and compression. At the Receiver, we solved the issue of lack of power for video processing by taking off those tasks from the main processor and putting them into the video card. By means of using new relatively 3D technologies like Pixel and Vertex Shaders, we transform a plain video image to the required especial display format, in real time without using much of the main CPU processing power.

The system provides a natural interface without the complications of 3D glasses etc. It is important to have a natural interface and that using the system does not hinder the experience and does not degrades compared to direct interaction, for instance, observing a fruit all its characteristics can be reproducible or be even better understood when observing the object through the stereo video system.

We presented then, a new tool for performing remote consultation, enabling a more effective plant management in fruit and vegetable production. By means of using stereo video, we can transmit images that reflect more reliably and realistically the characteristics of the observed object, allowing us to perform an efficient Remote Technological Assistance.

Future Work

We are currently working on the network component of the system; we are searching for an appropriate way of providing assured image quality, like for example Left-Right synchronization, to the transmission of stereo video, by means of network congestion control schemes and error recovery mechanisms.

Next stage involves taking the sender component of the system in order to stream directly from the actual scene where crops must be evaluated.

ACKNOWLEDGEMENTS

We thank the people at Sharp Corporation for providing us with a prototype of their 3D Display.

REFERENCES

[1] Iguchi, N., Uchio, F. and Takahashi, T. 2001. A Video Communication Method using Layered Coding and Packet Priority for Remote Technical Consultation on the Internet. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing.

[2] Johanson, M. 2001. Stereoscopic Video Transmission over the Internet. Proceedings of the Second IEEE Workshop on Internet Applications (WIAPP'01).

[3] Konrad, J. 1999. Enhancement of viewer comfort in stereoscopic viewing: parallax adjustment. Proc. SPIE Stereoscopic Displays and Virtual Reality Systems. 3639:179-190

[4] Ogawa, A., Kobayashi, K., Sugiura, K., Nakamura, O. and Murai, J. 2000. Design and Implementation of DV based video over RTP. Packet Video Workshop.

[5] Sierra, R., Iguchi, N., Uchio, F. 2003. A Time Limit and Hop Based Queue Management and Scheduling Technique for a Remote Technical Consultation System on the Internet. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing.

[6] Microsoft Corp. 2005. http://www.microsoft.com/windows/directx/default.aspx

[7] Sharp Corp. 2004. http://www.sharp3d.com

[8] Nvidia Corp. 2005. http://developer.nvidia.com/page/documentation.html

TABLES

Table 1. Performance gain when using GPU over CPU for one video frame processing.

	CPU (P4 2.4GHz)	GPU (GeForce 5700)	Gain
1-frame processing time	31 ms	9 ms	
Resultant Frame Rate	32 fps	111 fps	350 %

FIGURES

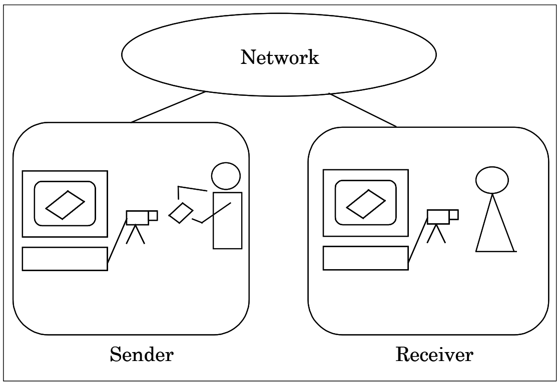

Fig. 1. Model of Remote Consultation

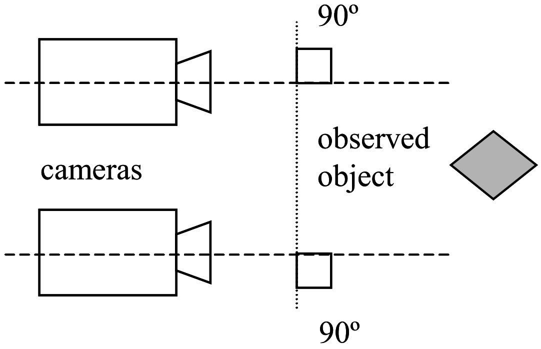

Fig. 2. Placement of cameras and object


Fig. 3. Mandarin Orange placed for remote stereo observation, with the pair of stereo cameras pointing to the objective


Fig. 4. Observed Mandarin Orange in expert's Sharp 3D Display

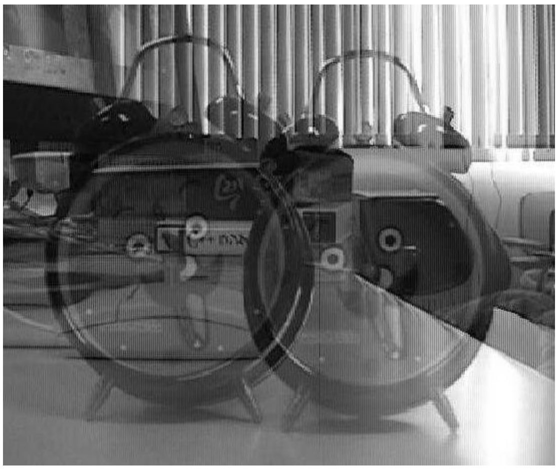
Fig. 5. Left and Right images overlapped, as taken by the cameras


Fig. 6. Left and Right images shifted to match the object on focus