Reproduction of Sound Signal from Gramophone Records using 3D Scene Reconstruction

Baozhong Tian and John L. Barron
Department of Computer Science
The University of Western Ontario
London, Ontario, Canada, N6A 5B7
{btian, barron}@csd.uwo.ca

Abstract

Preserving invaluable historic recordings has drawn some interest because the traditional record playback system wears out the record gradually. This paper presents a non-contact method to reproduce sound signal from gramophone records using 3D scene reconstruction of the micro-grooves cast on a record surface. Because of the unique shape of the microgroove, a planar assumption was made during scene reconstruction and recovered surface orientation was used to reproduce the sound signal. A robust estimation method was developed to reduce noise effects. Results from synthetic data were shown to test the technique.

Keywords: Sound Signal Reproduction, Scene Reconstruction, Gramophone Record, Robust Estimation, Surface Orientation.

1 Introduction

Reproducing sound mechanically on a record started as early as 1885 and the technology of recording and retrieving acoustic signals on gramophone records reached its peak during the 1970's, just before the digital format compact disk (CD) took over the mass marketing of music. Although the audio quality of CD is judged to be very good by most people, some audiophiles believe that the sampling rate of a CD (44.1kHz) is not high enough to reproduce the rich musical information faithfully. Today there is still some high-end record playing equipment in production. However, no matter how great a system performs, it wears out the record gradually due to the physical contact between the stylus and the record groove.

There also exists a lot of historical recordings that need to be archived. The problem with these recordings is that they have become so fragile that they can not tolerate being played back using a traditional style turntable with a mechanical stylus. This problem motivates research on non-contact record playing systems.

1.1 Traditional Method of Sound Reproduction

We will take stereo gramophone records (stereo LP) as our example, since other formats of mechanical records are similar. During the record cutting procedure, the left and right channel signals control the speed of the cutting stylus at a +45/-45 lateral manner, i.e. a composition of two orthogonal speeds perpendicular to each other, while the record rotates at a constant speed. This is called modulation of the grooves. The movement of the stylus determines the slopes in the tangential direction of the groove walls. This record cutting method keeps the left and right groove walls' modulation independent from each other. When the play back stylus has a similar setup as the cutting stylus, stereo signals can be reproduced. The electrical signal outputs are proportional to the +45/-45 lateral speeds of the stylus while riding along the groove and modulated by the groove walls.

Figure 1a illustrates the top view of the movements of record and stylus. The stylus has a tangential speed V_T relative to the groove due to the record rotation. There are also left and right lateral movements of the stylus (V_L and V_R) in +45/-45 directions. Figure 1b shows a cross section view of the compound +45/-45 lateral movement of the stylus.

The major goal of sound reproduction is to track the groove walls as precisely as possible. The conventional method uses a diamond-tip stylus to run along the V-shaped groove by applying a certain tracking force on the stylus. The problem is that the stylus has some weight, so the tracking of a

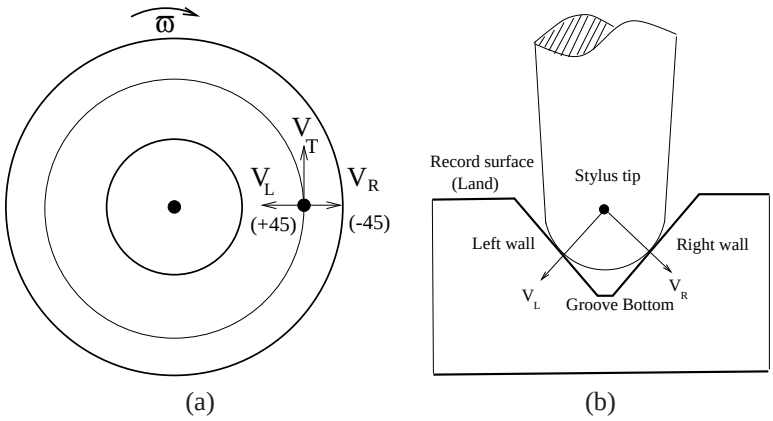


Figure 1: Illustrations of the movement of the record and the stylus: (a) top view, (b) cross section view.

high frequency signal is more difficult. The physical contact also produces a high temperature that softens the record surface and prevents it from playing well again within approximately 6 hours of time [Micrographia, 2005]. Other problems include groove damage such as scratches and small particles that result in annoying clicks, pops and degradation of sound over time and the maintenance of the correct settings of the turntable, tone arm, cartridge and stylus requires frequent adjustment.

1.2 Literature Survey of Non-Contact Record Playing Methods

Because of the problems with traditional record playback systems, people turned to the easy-to-use CDs as soon as they appeared in the early 1980s. But efforts at developing a non-contact record playback system did not cease. ELP corporation [ELP, 1997] spent ten years to develop a laser turntable (invented by Robert E. Stoddard et al. [Stoddard and Stark, 1989, Stoddard, 1989]) utilising five laser beams to track the microgroove optically. This is a pure analogue process, but it is so sensitive to foreign particles in the groove and on the record surface that it requires the record to be cleaned every time it is played. This ELP laser turntable uses two of the five beams of the laser to track the groove walls and the other three laser beams for groove tracking. This has two main advantages: the laser beams are weightless and can be made as thin as $2\mu m$ in diameter, which is much thinner than a high-end stylus (4-12μm). However, the system is very complicated and expensive and it only works well with black records because of the reflective nature of the material. Coloured records may produce unpredictable results [ELP, 1997].

Because the laser turntable is very expensive (in the price range of a small car) and because it is very sensitive to the cleanness of the record, some research has been carried out to study the feasibility of reproducing the sound signal by image processing methods. In 2002, Ofer Springer [Springer, 2002] proposed an idea he called the virtual gramophone. Springer's idea is to scan the record as an image and write a decoder to apply a "virtual needle" following the groove spiral form. However, when the authors listened to a sample decoded sound (http:www.cs.huji.ac.il/ ˜springer/), the music was judged to be barely recognisable. Inspired by Springer's idea, a group of Swedish students [Olsson et al., 2003] further developed a system to use more sophisticated digital signal processing methods such FIR Wiener filtering and spectral subtraction to reduce noise level in the reproduced sound, resulting in a better result than that of Springer's. Both systems used an off-shelf scanner, which limited the resolution of the images, to a maximum of 2400dpi or 10μm per pixel. At this resolution, the quantisation noise is quite high because the maximum lateral travel of the groove is about 150μm.

[Fadeyev and Haber, 2003] developed a 2D method to reconstruct mechanically recorded sound by image processing. The resolution was greatly improved by the aid of micro-photography. Their algorithm detects the groove bottom as an edge in the image and then differentiates the bottom edge shape to reproduce sound signals. Their method uses only the groove bottom information, which was not always very well defined and may be distorted by dirt particles. The groove walls, which contain rich sound informations, were ignored. They also introduced a 3D method to reproduce the vertically modulated records such as wax cylinders. But their 3D method requires complicated 3D profile scanning, such as that provided by a laser confocal scanning probe, which is a very slow process.

Johnsen et al. [Cavaglieri et al., 2001, Stotzer et al., 2003, Stotzer et al., 2004] also proposed a 2D method they called the VisualAudio concept. A picture of the record was taken using a large format film as big as the record. The film was then scanned using a rotating scanner, which is actually a line scan camera positioned above the film while the film is being rotated on a turntable. Edges were then

detected from the digitised image and then sound signals were computed from the edges. Unlike the method of [Fadeyev and Haber, 2003], Johnsen et al. used the groove and surface intersection as the edge instead of using groove bottom. This gives them the capability to reproduce the sound from stereo 33rpm recordings. Also the use of the rotating scanner eliminated the need for adjusting the sample rate as the groove turned close to the record's centre. The images are rectangular, and not circular, as scanned by a flat-bed scanner. A $10\times$ magnifier was fitted to the rotating scanner to get the desired image resolution. A Signal to Noise Ratio (SNR) analysis showed that a satisfying SNR of 40dB can be achieved if the standard deviation (σ_n) of edge position noise was kept below 1.28 μm. However, listening to the reproduced sound clips from their web site (http://www.eif.ch/visualaudio/) indicated that the noise level needs to be further reduced.

2 Proposed Method

We propose a sound reproduction method based on Computer Vision technologies such as optical flow and surface reconstruction. The proposed method uses a microscope to obtain a sequence of magnified images of the groove walls and uses 3D scene surface reconstruction to calculate the slopes of the walls. Figure 2 shows the system diagram. The major features of the proposed method can be summarised as:

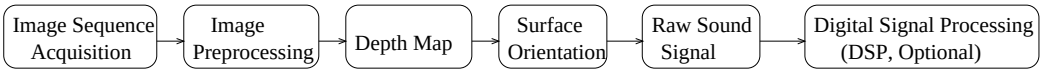


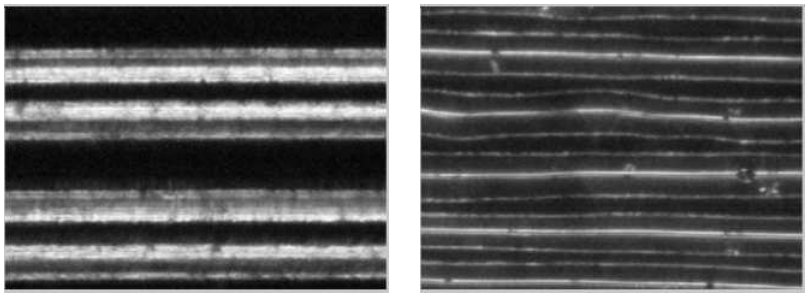
Figure 2: System diagram.

- Using as much information on the record as possible to reproduce the sound. Plenty of information is stored via the surface orientation of the groove walls, which is not used by 2D methods during their scanning/photographing processes. A 2D method only computes detectable edges such as a groove's bottom or groove-surface (land) intersections.
- Computer Vision technologies such as optical flow and depth map estimation are applied to this problem to obtain the 3D information characterising the groove, thus eliminating the requirement for a specialised 3D scanning device.
- Robust estimation techniques help choose the best areas of the groove wall for the computation and reject noisy areas which have been damaged by scratches and dirt particles, reducing the level of the noise and improving the quality of the reproduced sound.

We will discuss the individual system components below.

2.1 Image Sequence Acquisition

We need many groups of image sequence to cover the entire groove. Each group contains 36 frames of images. The two consecutive frames in same group should differ by only a few pixels. When the camera moves to the next segment of groove to capture another group of images, the last image of the current group should overlap with the last image of the previous group by a small amount so that the reconstructed groove is continuous when paired together.



(a) (b)

Figure 3: Two pieces of groove from (a) a 78rpm SP record and (b) a 33rpm LP record. The magnifying factor is 60X.

Figure 3 shows images of grooves under a microscope. The magnification factor of the microscope is set to be such that the field of view covers about $600\mu m$ in width so that for a camera with 640×480 pixels, the horizontal spatial resolution is about $1\mu m$. The illumination is set to a +45/-45 degree so that the groove walls are bright while the record surface and the groove bottom are dark. We are currently acquiring a better microscope with higher resolution and magnification camera to improve the image quality.

2.2 Image Preprocessing

The images have to be preprocessed, i.e. we need to compute the image intensity derivatives and the optical flow fields, etc. before we can compute depth maps. We experiment with implementations of two standard differential optical flow techniques, namely those of [Horn and Schunck, 1981] and [Lucas and Kanade, 1981] with differentiation by [Simoncelli, 1994].

The spatial-temporal differentiation method requires the scene to be rigid and smooth so that differentiation by convolution can be performed. Once differentiation has been performed, optical flow can be computed. Lucas and Kanade assume the flow is locally constant and use the motion constraint equation in a local least squares calculation to recover the optical flow. Horn and Schunck combined the motion constraint equation with a global smoothness term (the optical flow varies smoothly everywhere) to constrain the estimated velocity field in a regularisation (iterative) framework. We have to adapt these algorithms to our problem by imposing various constraint that arise from computing optical flow from record groove images, i.e. local surface planarity and uni-directional flow. The flow fields of Horn and Schunck look denser than that of Lucas and Kanade's but take a much longer time to compute. The computation and visualisation of depth requires a large number of such flow fields.

2.3 Depth Map Computation

We did a survey [Tian and Barron, 2005] of 4 recent algorithms for dense depth maps (from image velocities or intensity derivatives) which appeared to give good results in the literature. All of these algorithms assume known camera translation and rotation (or can be made to have this assumption). The 4 algorithms are those by [Heel, 1990], [Matthies et al., 1989], [Hung and Ho, 1999] and [Barron et al., 2003]. For a detailed description and discussions on these algorithms, please refer to [Tian and Barron, 2005]. Quantitative results show that the methods of Barron et al. is the best over all.

We report here experimental results for Barron et al.'s algorithm on synthetic record groove images and on real groove images with encouraging results. Because the groove wall orientation can be described by 2 angles, one of which is constrained and because the vertical component of image velocity is always very small (uni-direction constraint), we anticipate imposing such constraints will yield better even results. For example Barron et al.'s method could be modified to use only horizontal velocities, like Matthies et al. and effectively have only one angle of the surface orientation to track in the Kalman filter.

2.4 Robust Estimation of Surface Orientation

Surface orientation is computed from depth using least squares. Assuming local planarity, the surface orientation $\hat{\alpha}$ of a local neighbourhood is constant and satisfies the planar equation $\hat{\alpha}\cdot\vec{P}=c$, where $\vec{P}=[X,Y,Z]$ is the 3D coordinate of a pixel and c is a constant. We can solve this linear system using a robust estimation method called Local M-Estimates, recommended in Press et al. [Press et al., 1992]. This should give more accurate surface orientation as outlier data will be suppressed.

We present a robust estimation formulation of this calculation as follows. We use vector $\vec{g}=(g_1,g_2,g_3)$ to denote $\frac{\hat{\alpha}}{c}=(\frac{\alpha_x}{c},\frac{\alpha_y}{c},\frac{\alpha_z}{c})$. We can set up a least square system:

$$W\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \vdots & \vdots & \vdots \\ a_{N1} & a_{N2} & a_{N3} \end{bmatrix}\begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} = W\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}, \tag{1}$$

or

$$WA\vec{g}=WB \tag{2}$$

where W is a $N\times N$ diagonal matrix with diagonal elements acting as the weights for the N equations, and

$$a_{i1} \quad = \quad X_i \tag{3}$$

$$a_{i2} = Y_i \quad (4)$$
$$a_{i3} = Z_i \text{ and} \quad (5)$$
$$b_i = 1. \quad (6)$$

The solution is $\vec{g} = (A^T W^2 A)^{-1} A^T W^2 B$.

The weight matrix W plays a critical role in this robust estimation calculation. Initially, W is I so that a rough solution is obtained. Using this solution, we can refine W using the Lorentzian estimator, [Black and Anandan, 1996] ρ:

$$\rho(d_i, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{d_i}{\sigma}\right)^2\right) \quad (7)$$

and the influence function ψ (which is the derivative of ρ):

$$\psi(d_i, \sigma) = \frac{2d_i}{2\sigma^2 + d_i^2}, \quad (8)$$

where σ is a scale parameter and d_i is the residual value of each equation:

$$d_i = |a_{i1}g_1 + a_{i2}g_2 + a_{i3}g_3 - b_i|. \quad (9)$$

Then the weight matrix elements get updated as:

$$w_i = \frac{\psi(d_i, \sigma)}{d_i}. \quad (10)$$

We can re-calculate $\vec{g}$ again using the updated W. This procedure is repeated until one of the following stopping criteria is met:

- the total residual is smaller than some threshold: $||d_i||_2 < \tau_1$,
- the total residual begins to diverge: $||d_i||_{t^-} - ||d_i||_t < \tau_2$ or
- the number of iterations reaches a limit.

The second threshold, τ_2, which is a small positive number, allows the total residual to vary up and down a bit before iterations are considered to be converging or diverging.

According to Black and Anandan, tuning the scale parameter σ may work well given that the initial approximation for it is not too bad. Since in the Lorentzian estimator, a residual d_i is considered an outlier if $d_i \geq \sqrt{2}\sigma$, lowing σ after each iteration will reveal more and more outliers. Another benefit we can get from this is that the number of outliers could help us to determine whether the value of σ is properly chosen and when to terminate the iterations.

2.5 From Surface Orientation to Sound Signal

Once the surface orientations are computed, they need to be interpreted into sound signals so that they can be played. Figure 4 illustrates a piece of a groove, showing the left surface orientation $\hat{\alpha}_L$. The figure also shows the two angles (θ_{XY} and θ_{YZ}) that determine $\hat{\alpha}_L$. Due to the +45/-45 modulation of the groove walls, θ_{YZ} is approximately 45 degrees at all times. Note also that locally the surfaces are planar. To extract the left channel signal, we observe that the surface orientation lies in the plane $z = y$ because of the +45/-45 stereo modulation. Accordingly, the surface orientation corresponding to the right channel lies in the plane $z = -y$.

We define θ_L to be the modulation angle between the surface orientation $\hat{\alpha}$ and the left-channel-zero-modulation orientation $\hat{n}_L = [0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]$, which is the surface orientation of the left groove wall when the signal is zero. Then the ratio of the lateral speed V_L and the tangential speed V_T of the stylus is:

$$\frac{V_L}{V_T} = \tan\theta_L \quad (11)$$

where

$$\theta_L = \arccos(\hat{\alpha}_L \cdot \hat{n}_L). \quad (12)$$

V_L corresponds to the left channel signal and needs to be adjusted according to $V_T = \omega R$, where R is the current distance to the record centre. A similar method can be applied to reproduce the right channel

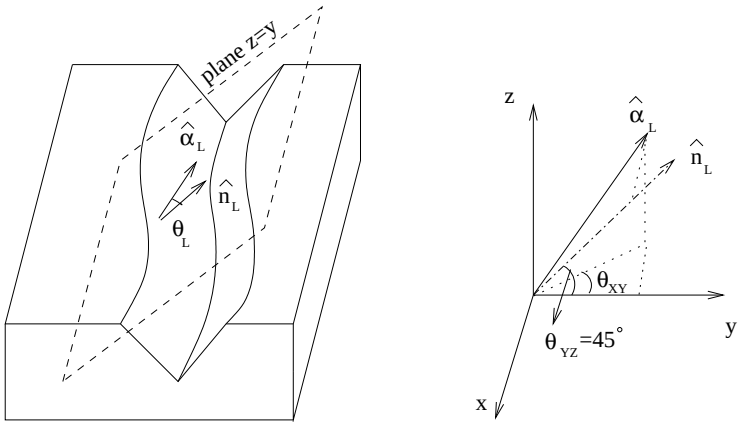


Figure 4: Illustration of a piece of groove showing the surface orientation $\hat{\alpha}_L$ of the left groove wall lies in the plane of $z = y$.

signal, V_R, using the direction $\hat{n}_R = [0, -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]$ instead of $\hat{n}_L$. A mono recorded gramophone record with a horizontal groove modulation can be treated as a special case of the stereo groove modulation, where $V_L = V_R = V$, so either the left or right surface orientation can be used to reproduce the sound signal.

For a mono SP or a wax cylinder with vertical groove modulation, we can project the surface orientation onto the vertical $x - z$ plane, i.e. $y = 0$, and θ and V can be calculated using above equations, except now:

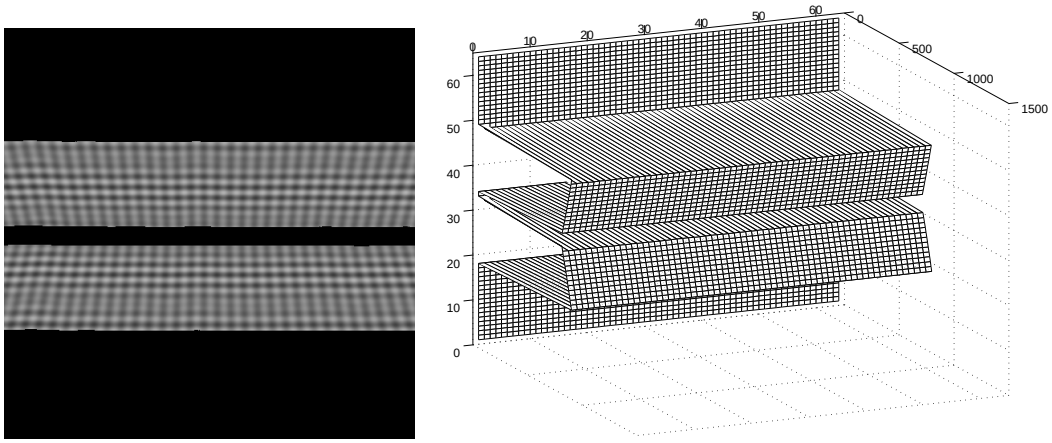
$$\hat{\alpha} = [\alpha_x, 0, \alpha_z] \text{ and} \tag{13}$$

$$\hat{n} = [0, 0, 1]. \tag{14}$$

Due to the robustness of the algorithm introduced in section 2.4, we anticipate that the algorithm will be able to reject most of the noise such as pops, clicks caused by scratches or small dirt particles, etc. However, comparing the waves in Fig 7b and Fig 7c indicate that post-processing may be needed to reduce the noise.

3 Simulation Technique

Implementation of our technique with real record data is currently underway. In this section, we report experimental results with synthetic data. We generated groups of ray-traced groove image sequences

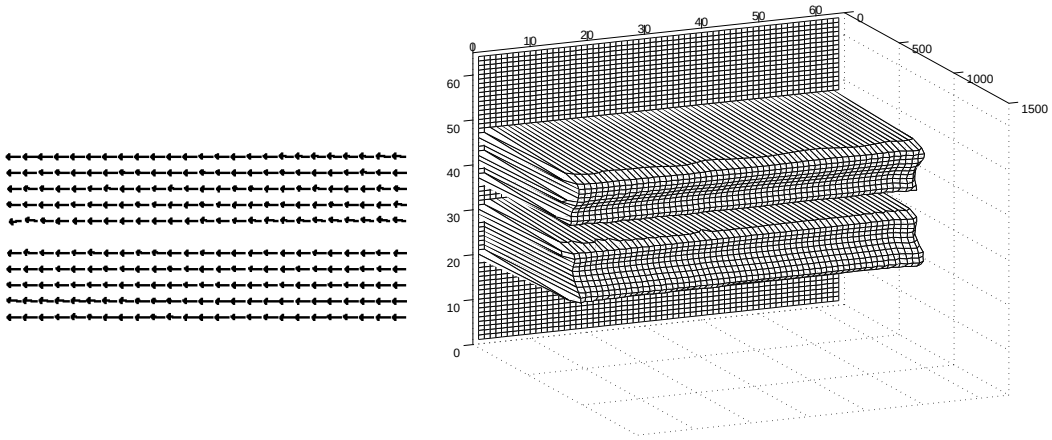


(a) (b)

Figure 5: Synthetic test data: (a) A sinusoid-texture groove (b) The 3D perspective true depth map of the groove.

with the camera translating to the left by $(1, 0, 0)$, an example of which is shown in Figure 5. The offset

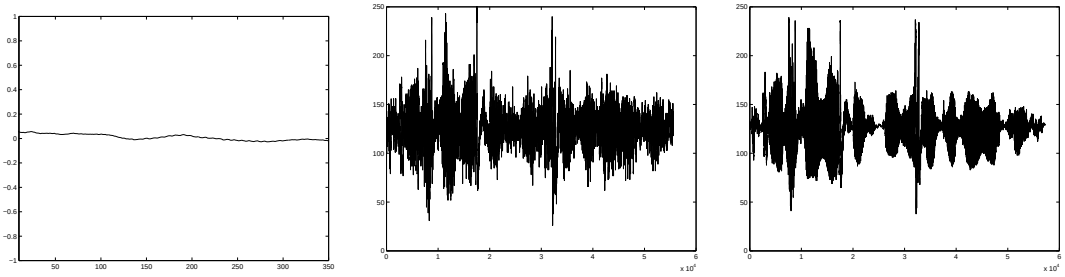
of groove walls are modulated by a man's voice. For about 2 second clip ("Computers are useless. They only give you answers" - Pablo Picasso) we generated 1390 groups of such images with each group having 36 images. From each group, we can recover a piece of the groove depth map and, hence, a piece of the sound. By 'stitching' these small pieces together, we obtain the complete sound recording. Optical flow was computed from these sequences of images as shown in Figure 6a. In this experiment,



(a) (b)

Figure 6: Test results: (a) Optical flow computed from the synthetic image sequences. (b) The recovered depth map of the groove.

we used Horn and Schunck's algorithm to obtain a smooth dense flow field. Next, we fed this flow sequence to Barron et al's depth recovery algorithm, which incorporates a Kalman filter to compute a smooth surface reconstruction. The recovered depth map is shown in Figure 6b.



(a) (b) (c)

Figure 7: Sound waves: (a) The recovered sound wave from one piece of synthesised groove, (b) The recovered sound wave from groups of such images and (c) The original sound wave.

After the depth maps were computed, surface orientations of the grooves were then estimated. From the surface orientations, sound signals were computed as introduced in 2.5. The sound wave pieces such as those shown in Figure 7a were combined together to form the total sound wave as shown in Figure 7b. Compared with the original sound wave shown in Figure 7c, the reconstructed sound is very similar to the true sound, although there is some noise present, as can be seen in Figure 7b. We compute the shape envelopes of the computed and the original waveform and then compute Pearson's product-moment correlation coefficient of them as $r = 0.848$ which indicates good correlation. Listening test confirms (available at www.csd.uwo.ca/~btian/IMVIP2006) that the sound is recognisable in spite of the presence of noise. Further refinements in the algorithms and the use of higher resolution images may be able to attenuate the noisy components of the retrieved sound.

4 Conclusions

This paper established a framework for recovering sound from gramophone records through 3D reconstruction. This may not necessarily be a real-time system due to such limiting factors as the computation cost and camera speed, (although we believe technology advances will eventually overcome these limitations). Our algorithm has the potential of recovering sound from damaged records such as scratched

or even broken records. We are investigating the feasibility of a real-time sound reproduction system, such as hardware implementations of the preprocessing steps, fast image acquisition, etc.

References

[Barron et al., 2003] Barron, J. L., Ngai, W. K. J., and Spies, H. (2003). Quantitative depth recovery from time-varying optical flow in a kalman filter framework. In T. Asano, R. K. and Ronse, C., editors, *LNCS 2616 Theoretical Foundations of Computer Vision: Geometry, Morphology, and Computational Imaging*, pages 344–355.

[Black and Anandan, 1996] Black, M. J. and Anandan, P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104.

[Cavaglieri et al., 2001] Cavaglieri, S., Johnsen, O., and Bapst, F. (2001). Optical retrieval and storage of analog sound recordings. In *The AES 20th International Conference*, Budapest, Hungary.

[ELP, 1997] ELP (1997). Elp laser turntable. Internet reference: www.elpj.com.

[Fadeyev and Haber, 2003] Fadeyev, V. and Haber, C. (2003). Reconstruction of mechanically recorded sound by image processing. *J. of Audio Eng. Soc.*, 51(12):1172–1185.

[Heel, 1990] Heel, J. (1990). Direct dynamic motion vision. In *Proc IEEE Conf. on Robot Automation*.

[Horn and Schunck, 1981] Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17:185–204.

[Hung and Ho, 1999] Hung, Y. S. and Ho, H. T. (1999). A kalman filter approach to direct depth estimation incorporating surface structure. *IEEE PAMI*, pages 570–576.

[Lucas and Kanade, 1981] Lucas, B. D. and Kanade, T. (1981). An iterative image-registration technique with an application to stereo vision. In *Image Understanding Workshop*, pages 121–130. DARPA.

[Matthies et al., 1989] Matthies, L., Szeliski, R., and Kanade, T. (1989). Kalman filter-based algorithms for estimating depth from image sequences. *IJCV*, 3(3):209–238.

[Micrographia, 2005] Micrographia (2005). The microscopy of vinyl recordings. Internet reference, www.micrographia.com.

[Olsson et al., 2003] Olsson, P., Öhlin. R. Olofsson, D., Vaerlien, R., and Ayrault, C. (2003). The digital needle project - group light blue. Technical report, KTH Royal Institute of Technology, Stockholm, Sweden. Internet resource: www.s3.kth.se/signal/edu/projekt/students/03/lightblue/.

[Press et al., 1992] Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1992). *Numerical Recipes in C*. Cambridge University Press, 2 edition.

[Simoncelli, 1994] Simoncelli, E. P. (1994). Design of multi-dimensional derivative filters. In *IEEE Int. Conf. Image Processing*, volume 1, pages 790–793.

[Springer, 2002] Springer, O. (2002). Digital needle - a virtual gramophone. Internet resource: www.cs.huji.ac.il/˜springer/.

[Stoddard, 1989] Stoddard, R. E. (1989). Optical turntable system with reflected spot position detection. United States Patent 4,870,631.

[Stoddard and Stark, 1989] Stoddard, R. E. and Stark, R. N. (1989). Dual beam optical turntable. United States Patent 4,870,631.

[Stotzer et al., 2003] Stotzer, S., Johnsen, O., Bapst, F., Milan, C., Sudan, C., Cavaglieri, S. S., and Pellizzari, P. (2003). Visualaudio: an optical technique to save the sound of phonographic records. *IASA Journal*, pages 38–47.

[Stotzer et al., 2004] Stotzer, S., Johnsen, O., Bapst, F., Sudan, C., and Ingol, R. (2004). Phonographic sound extraction using image and signal processing. In *Proc. ICASSP*, Montreal, Quebec, Canada.

[Tian and Barron, 2005] Tian, B. and Barron, J. L. (2005). A quantitative comparison of 4 algorithms for recovering dense accurate depth. In *2nd Canadian Conference on Computer and Robot Vision*, pages 498–505, Victoria, BC, Canada.