UCam: Direct Manipulation using Handheld Camera for 3D Gesture interaction

Liang Zhang
Department of Computer Science and Technology
Tsinghua University
China
blinkzhangpaul@gmail.com

Yuanchun Shi
Department of Computer Science and Technology
Tsinghua University
China
shiyc@tsinghua.edu.cn

Mingming Fan
Department of Computer Science and Technology
Tsinghua University
China
fmmbupt@gmail.com

ABSTRACT
This paper presents UCam a novel approach in 3D Gesture Interaction based on handheld camera movement. UCam reflects hand's movement and directly maps it to the movement of 3D object based on visual tracking of feature-like points on incoming frames. Only one button is needed to differentiate rotation and translation. The advantages of this technique lie in the popularity and low cost of handheld cameras, low requirement and no need of adjustment of background and easy to use for beginners. To evaluate UCam, it is compared with mouse in some 3D controlling tasks. The results show that UCam is more flexible and easier to use and master in most cases. Even for complicated tasks, UCam has comparable performance as mouse.

Categories and Subject Descriptors
H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

General Terms:
Human Factors

Keywords
Interaction, Camera Movement, Direct Manipulating

1. INTRODUCTION
Due to the complexity of 3D interaction's combination input devices and operation method, which requires 6DoFs, users are kept at a respectful distance. However 3D Gesture Interaction, which is totally based on hand's gesture, can directly manipulate 6DoFs. As a result, it offers easier induction to the technique for users.

In this paper, 3D interaction means the interaction in the space, which requires distinct degrees of freedom (6DoFs)—translating, zooming and rotation by x,y,z axes.

It is usually regarded that 3D movement, which contains only translating and zooming, constitutes a part of 3D interaction. In this paper, we define the term translating as translating by x, y axes on 2D plane and the term zooming as translating by z axis.

Because of its familiarity to users and easy to manipulate, handheld cameras is suitable for 3D gesture interaction. Common handheld cameras are able to directly reflect the hands' gestures and map them into the 3D interaction. Mobile phone with camera could be a supporting device for its existing platform and application. Mobile service based on camera adaptive viewing [1] [2] offers many possibilities of application on mobile phones with cameras.

UCam is an interaction technique which has low learning effort for its direct mapping of hand's movement in 6DoF space to the movement of a 3D object. Based on tracking feature points, e.g. corner points on visual background, from incoming frames, UCam is able to differentiate at most 4DoFs of camera's movement including translating, zooming and rotation by z axis or rotation by x, y, z axes and zooming. To support the other two degrees, we use a button control (see Figure 1). As a 2D input device instead of 3D movement, handheld camera offers tremendous convenience for users in 3D gesture interaction.

2. RELATED WORK
Handheld cameras have been explored as input devices in many mobile applications [1][2][5]. Panu Vartiainen[5] presented a mobile visual interaction system enables pointing with mobile camera devices on large displays based on identifying and interpreting camera-detectable data elements. With the system, users need to point the camera directly towards the specified displays. On the contrary, UCam supports the common use of camera's interaction everywhere based on tracking corner-like features on incoming frames.

Antonio and Koichi [1] proposed a tracking algorithm that applies the 2D movement of some tracked points in applications such as picture browsing, document viewing, etc. Jingtao Wang [2] proposed *TinyMotion as* a pure software approach to detect hand's 2D movement, which is evaluated using camera phone as a handwriting capture device. The limitation of these applications is that they could only handle 2D hand's movement.

Several techniques have been proposed to achieve 3D interactions based on handheld camera. Adel and Peter [3] designed an algorithm to automatically generate the camera path and map the path into 3D space navigation. However, it is limited to a fixed visible background. Moreover, in their method, initialization and adjustment of light on background is required. In contrast, UCam

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.

does not require any specific and fixed background. As a result, Ucam is more robust and flexible.

Unicam [4] is the most similar technique to Ucam. However, instead of requiring supporting devices such as a single-button stylus or a mouse to invoke a single 3D view, UCam can differentiate independently at most 4DoF movement of hand—2D translating, zooming and rotation by z — simply by using a button to transform translating to rotation by x, y.

Figure 1. Picture 1 shows the platform of experiment. Picture 2 uses UCam to view a 3D object in large display. Picture 3 shows the handheld camera used in UCam.

3. DESIGN

3.1 Tracking Alogrithm

UCam uses corner points which could be effectively detected, easily tracked and relatively stable. UCam has no initialization phrase and little requirement of the background. However, to make the best of UCam, blankness or single color background and fast moving objects are not recommended (see Figure 1).

Based on the change of the 2D coordinates of corner points, camera's movement could be differentiated and calculated (see Figure 2).

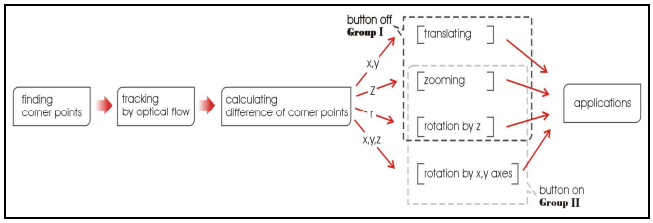

Figure 2. The working process of UCam. A button is used to divide the movements into two groups. The quantity of each movement will be applied in 3D interactions.

3.2 Specification and Calculation

6DoF movements include translating, zooming, and rotation by x, y, z axes (see Figure 3). UCam can at most specify 4D oF, because each movement invokes an independent change for corner points. Supplied with a single button, we can differentiate translating and rotation.

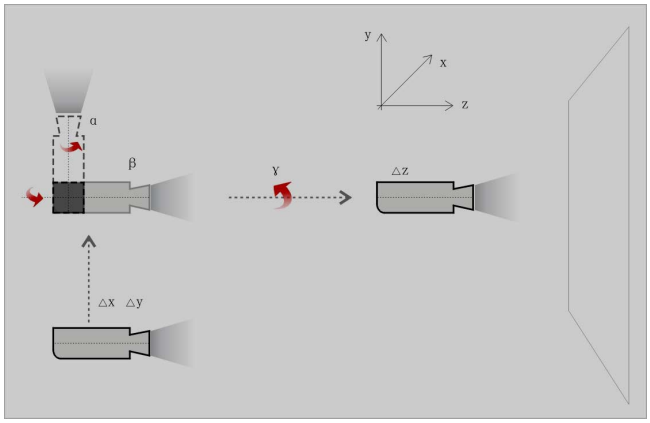

Figure 3. We denote 2D coordinates of translating as (x,y) , zooming rate as z, rotation degree by x,y,z axes as α, β, γ .

Translating is calculated by the average change of corner points' coordinates from two neighboring frames. The influence of zooming and rotation by z on coordinates should be excluded

Zooming rate is reflected by the change of the distance between corner points and the center point (see Figure 4).

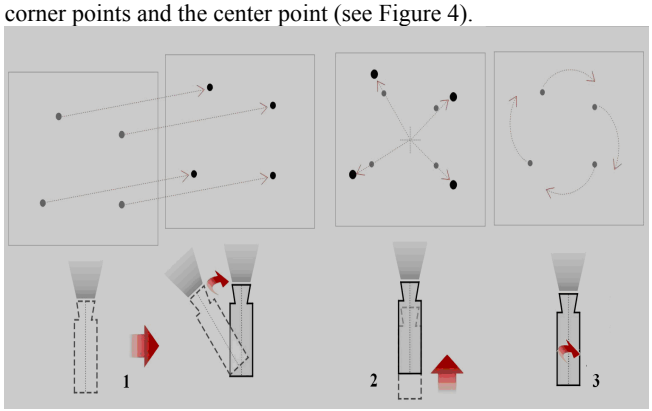

Figure 4. Picture 1 shows translating and rotation by x,y both result a same changing directions for all the points. Picture 2 shows points radiate or shrink from the center point in zooming. Picture 3 shows points in rotation by z axis go in a clockwise or anticlockwise direction.

Since the rotation by x, y can be simulated by the combination of zooming and translating (see Figure 4), the quantity of translating and constants adjusted to the zooming rate could be used to calculate the rotating degree.

The degree of rotation by z can be figured out from the change of slopes of lines between corner points and the center point.

Although the above calculation is not very accurate, it is efficient to differentiate those movements with satisfying performance in 3D interaction.

4. EVALUATION

3D tasks are used to evaluate UCam. Users are required to control a 3D cube rendered by OpenGL by a handheld camera and fill it in 6 locations. Those locations involve different quantity of translating, zooming rate, rotation degree with difficulty gradually increased from 1 to 6. We use a handheld camera which can grab 30 frames per second of 320*240 pictures. The background is focused on the surface of a big table with books and files on it, which can be considered as a stable and common background for tracking algorithm (see Figure 1).

We compare the efficiency of mouse and camera by the time being used to accomplish these tasks. Users are specified as 4 groups—skilled in using camera, skilled in mouse, unskilled in mouse and unskilled in mouse. Users skilled in mouse are defined as students who used to mouse in 3D games while users skilled in camera are those who use UCam above an hour everyday for a month. And we selected people who rarely use controlling devices as unskilled in both mouse and camera. 10 people for each group are tested in this experiment. All the users were given 20 minutes to get familiar with the controlling commands in these tasks. The time to accomplish the 6 tasks was recorded.

Table 1. Six locations applied in experiment involving a series values of x, y, z, α, β, γ. Difficulty gradually increases from location 1 to 6.

No.	x	y	z	α	β	γ
1	1.5	1.5	1.0	0	0	0
2	-2.0	-1.5	1.5	0	0	0
3	2.0	-1.5	0.8	0	0	15^0
4	2.5	1.0	0.5	0	30^0	30^0
5	-2.5	1.0	1.2	45^0	-65^0	105^0
6	-1.5	-2.2	2.0	-60^0	15^0	220^0

In this experiment, we recommend the users to adjust the 3D view of the cube at first and then fill it into the locations.

4.1 Easy Learning

The first two curves in Figure 5 indicate that there is little difference with users skilled and unskilled in camera but large difference between those using mouse, which means handheld camera is much easier to learn than mouse as an approach of interaction. In fact, mapping 2D navigation of mouse with buttons and wheel to 3D movement is not as intuitionistic in human's mind as their hands' real 6DoF movement is. They need more time to learn and get familiar with complicated interacting commands of mouse in 3D environment. However, UCam ,which reflects directly hand's 3D movement, is more understandable for fresh users who rarely use computer devices.

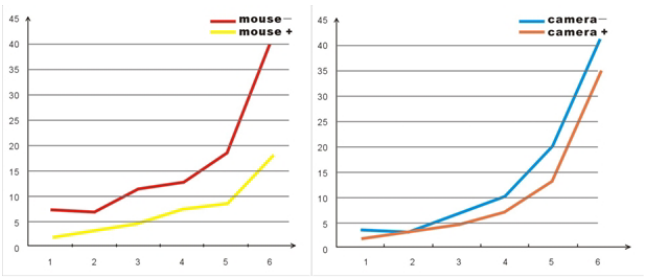

Figure 5. '-' means unskilled users and '+' means skilled users. The figure compares the time of each two groups of users to accomplish tasks. Task 6 is of the highest difficulty and task 1 is of the lowest difficulty.

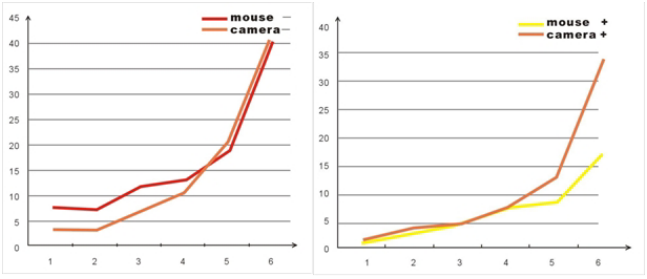

Figure 6. '-' means unskilled users and '+' means skilled users. The figure compares the time of each two groups of users to accomplish tasks. Task 6 is of the highest difficulty and task 1 is of the lowest difficulty.

Users unskilled in both mouse and camera turned out differently in the tasks (see Figure 6). Mouse users took more time to accomplish tasks 1,2,3,4 than camera users and approximately the same time in tasks 5, 6 as camera users. This result also identifies the former conclusion that UCam reflects direct hand's 3D movement rather than mouse's complex interacting commands.

4.2 Efficiency in Interactions

However, when it comes to users skilled both in mouse and camera, the camera acts same efficiently as mouse does in tasks 1,2,3,4 but more difficultly in tasks 5, 6.(see Figure 6) Task 1,2 involves only translating and zooming, task 3,4 involve simple rotation and task 5, 6 involve the sophisticated 6DoF.

Rotation is more limited than translating and zooming on hands for human's physical limitation. It is impossible for a user's hand to achieve a large degree of rotation (larger than 90 degree) owing to the limitation of the wrist. In such circumstances, users' rapid return of camera though has been declined in the algorithm, still cause inaccuracy in controlling.

In addition, provided no support plane, total freely controlling a camera in 3D space causes more dither than mouse on a 2D controlling plane. Therefore, due to the dither and limitation, users often need more time to adjust the degree in rotation by camera than mouse.

Furthermore, in our algorithm, fast transmitting background will cause problems in the consistency of movement's calculation. A sudden movement from a near background to a far background will result in a jump of calculation or loss of corner points, which will result in instability. It often leads to overdue adjustment in rotation for common users.

The proportion between visual movement captured by camera and the cube is far from perfect to achieve accuracy. It is hard to evaluate the proportion and need to optimize in further practices.

4.3 User Experience

From the users' experience for this new interaction, it can be figured out that camera is more instable and less accurate than mouse in sophisticated 6DoF movement, but much easier to learn, more flexible and more convenient than mouse (see Figure 7).

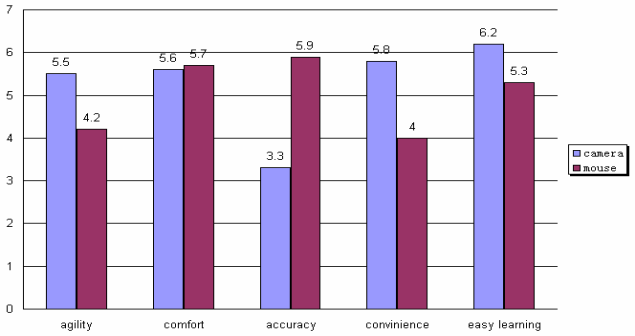

Figure 7. Users' experience on using camera and mouse in our interacting tasks. 7 is the highest score, and 0 is the lowest.

As a 2D input device, mouse has to use two buttons and a wheel in 3D interaction, which causes sophisticated controlling commands. However, UCam which requires only a button simply takes hand's movements as basic controlling commands. Foregoing reasons contribute to UCam's easier learning process. Moreover, portability and low requirement for background of UCam make it more convenient and flexible.

4.4 Division of Movements

Furthermore, users thought the division of two groups---translating, zooming, rotation by z and rotation by x, y, z ,zooming, is more convenient compared with mouse's complex controlling---buttons and wheels. In most of existing PC games, players need to turn the object towards a certain direction (rotation by z), then navigate on a 2D plane (translating) and a zooming view of this object is also required. Group I of movements can totally satisfy the need for these games. In addition, group II can be applied to browse a static object in 3D view. Those two groups with a button as transition cooperate efficiently in our experiment.

5. CONCLUSION AND FUTURE WORK

This work shows the potential of handheld camera movement based on 3D gesture interaction that requires no fixed background and low effort to learn with instant feedback. This technique opens up a new application and interaction opportunities with public and private interaction such as Google earth, 3D games and also cover many 2D applications on mobile service[1][2].

In order to achieve more accuracy and precision in control and more stability in rotation, future work will concern the improvement in the algorithm to get 3D movement parameters of handheld camera. A multi-camera interaction might be an interesting topic as well.

6. REFERENCES

[1] Antonio Haro, Koichi Mori and Vidya Setlur, Tolga Capin：Mobile Camera-Based Adaptive Viewing. *In Proc. CHI 2005*, ACM Press (2005), 78 – 83.

[2] Jingtao Wang, Shumin Zhai and John Canny: Camera Phone Based Motion Sensing: Interaction Techniques, Applications and Performance Study. *In Proc.CHI 2006*, ACM Press (2006), 101 – 110.

[3] Adel Ahmed, Peter Eades:Automatic Camera Path Generation for Graph Navigation in 3D. *In Proc. CHI 2005*, ACM Press (2005), 27-32.

[4] Robert Zeleznik' Andrew Forsbergt: UniCam - 2D Gestural Camera Controls for 3D Environments. *In Proc. CHI 1999*, ACM Press (1999), 169-173.

[5] Panu Vartiainen, Suresh Chande, Kimmo Rämö.Mobile Visual Interaction Enhancing local communication and collaboration with visual interactions. *In Proc. CHI 2006*, ACM Press (2006).