The Video Cube Puzzle: On Investigating Temporal Coordination

Eric Yim, William Joseph Gaudet, Sid Fels

Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, B.C., Canada
kyim@ece.ubc.ca, joe@joegaudet.com, ssfels@ece.ubc.ca

Abstract. We have created a novel computer-based 3D puzzle, named *Video Cube Puzzle* to investigate human beings' temporal coordination abilities. Ten adult participants were studied solving ten cubic video puzzles of varying difficulties using a within-subject design. The ten puzzles have two segmentation variations, 2x2x2 and 3x3x3, and five texture variations, solid colours and four videos of drastically different contents. Only 60% of the subjects were able to complete the entire problem set. The results indicate that random imagery and “active” videos make for easier Video Cube Puzzles. Similarly, a geometric increase in difficulty was noted as the number of segments in the puzzle increased. The challenging nature of temporal video cube puzzles appears to be partly due to people’s poor ability to process temporal information when using a spatial representation of the timeline using a three dimensional volume. Additional studies are suggested to explore this further. As a new type of game however, the Video Cube Puzzle allows the complexity of the puzzle to be easily varied from simple to extremely complex providing a way to have a continuous pathway to skill and control leading to a satisfying experience when the puzzle is solved.

Keywords: Multimedia, Video Game, Video Cubism, tx-transform, Slit-scan, Puzzle, ePuzzle.

1 Introduction

Jigsaw puzzles have been around since the 18th century. The most common way we see them today is in the form of cut-out cardboard pieces that, when assembled in the correct way, will reproduce a given 2D image. In addition to taking different layouts in 2D, jigsaw puzzles have also taken many shapes and forms throughout its history – they can be categorized into the physical and virtual departments. In the physical department, there are, for example, double-sided jigsaws, puzzles showing optical illusions, 3D jigsaw puzzles that assemble to a castle and various other objects, and a subgroup of 3D jigsaws that are spherical and are also known as puzzle globes. The virtual department, with the help of computer processing, provides even richer variety. For instance, there are virtual puzzles that simply translate the physical ones

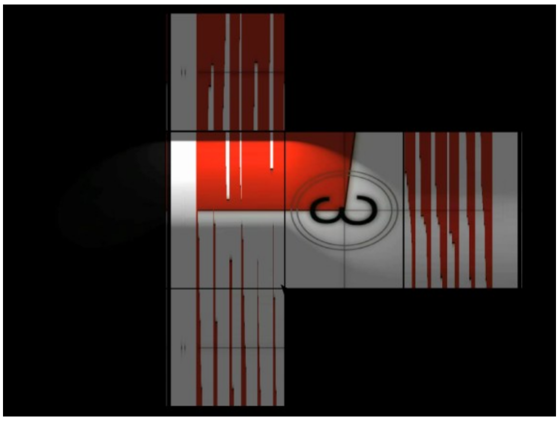

Fig. 1. The faces of a Video Cube, loaded with a countdown video, arranged in paper-model-like fashion. The usually "hidden" sides of the video become a source of interesting visual artifacts exploited by the Video Cube Puzzle.

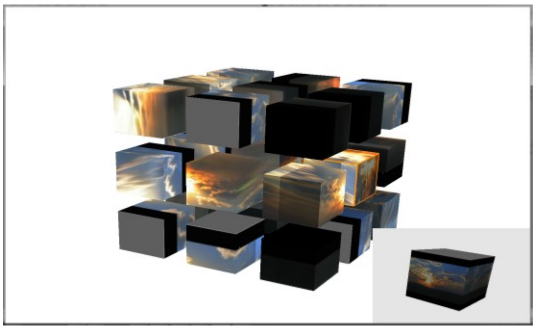

Fig. 2. Interface of the Video Cube Puzzle. The larger view in the middle shows the puzzle portion that the player interacts and solves, whereas the smaller view at the corner shows the reference that the player tries to match.

directly into a virtual environment, popular puzzle video games like Tetris[1] and the Puyo Puyo series[2] which are spin-offs of the central concept, and an audio version that splits up a soundtrack into scrambled segments and requires players to use their ears to solve the puzzle [1]. Despite the diversity in representations, however, the underlying concept has remained grossly unchanged – identify the spatial relationships between puzzle pieces and assemble or rearrange these pieces to achieve a visual goal. In other words, they mainly challenge our spatial coordination.

[1] Tetris is a popular puzzle video game that uses falling tetrominoes as its game pieces in a grid. The objectives are to manipulate and stack the falling game pieces to fill and erase as many horizontal lines as possible to collect points and to avoid stacking over the grid.

[2] Puyo Puyo is a popular Japanese falling pieces puzzler. The game is always played against at least one opponent, computer or human. The objective is to outlast your opponent by connecting as many sets of four same-coloured puyos as possible so that they can disappear from your grid and become obstacles on the opponent's grid.

Given the efficiency and effectiveness of the human race in solving such spatial riddles, the objective of this paper is to investigate and discover human beings' abilities to solve a novel virtual puzzle called *Video Cube Puzzle* (VCP), written using OpenSceneGraph (OSG) that has an added temporal dimension of difficulty.

A *Video Cube* is a virtual cube that hosts video frames in its volume rendering it a cubic video player. It is inspired by the work done on Interactive Video Cubism [7, 8]. The front and back faces mark the ends of the video. So, when seen from the front face, the cube appears as a movie screen displaying the first frame of the video. The video can be played on the cube. That is, a parallel, invisible cut-plane passes through the cube from front to back at a constant rate while also stepping through the video frames at the same rate. This way, the entire clip will appear as being played on the cube as it is being consumed. The edges of the video frames form the side images of the cube. These side images are non-existent when the video played in the traditional 2D fashion and, thus, are an interesting artifact of the Video Cube.

An example is shown in Fig. **1**. This Video Cube is loaded with a countdown video often seen in films, and its texture-mapped faces are arranged in a paper-model-like fashion for clarity. As clearly seen, all faces, except for the front and back, display strange images that one would not expect to see in a countdown video; yet, these side images are created from the often-ignored edge pixels from the same clip.

The puzzle is formed by segmenting the cube into smaller, equal-sized sub-cubes after which they are randomly displaced in both positions and orientations. Because the video plays only along one direction of the cube, a strange visual effect known as tx-transform [14] or slit-scan [12, 15] can be observed on disoriented sub-cubes. Players are asked to rearrange the sub-cubes to resemble the given solution much like the objective of the Rubik's Cube. But the similarity ends there because, unlike the Rubik's Cube, players must manipulate the cubes individually for this puzzle. The allowable operations are 90-degree rotations around the three principle axes and positions swapping between two cubes. Players are asked to take advantage of the continuum of the video source to solve the puzzle.

In addition to being an interesting artifact, the side images also provide temporal cues that aid in solving the VCP. Although in static form they appear as surface mapped texture that seems to carry only spatial information, these temporal-data-formed images are usually random, irregular, and confusing to a point that would not make sense unless put in a dynamic, temporal context. These temporal cues can effectively convey the subtle dynamic changes of the video clip.

Like traditional jigsaw puzzles, the difficulty of the VCP can be raised by increasing the number of segmentation. Moreover, difficulty can be further raised by choosing a video source with low spatial and/or temporal variations.

2 Related Work

As mentioned previously, there are many varieties in the world of virtual jigsaw puzzles. Many of these computer-based adaptations are considerably easier than their physical counterparts because they do not allow pieces to be rotated, thus reducing the number of degrees of freedom. For example, all the puzzles on the popular website

JigZone.com are in that nature. Contrarily, there is a popular kind of on-line jigsaw puzzle, called moving video jigsaw puzzle, that uses a constantly looping video as their image source and, as a result, makes for a more challenging and interesting experience [3]. This is an example of an integration of a video source with a puzzle.

Cube is a common element used in puzzles and games. Some of the interesting applications that use cubes as their main puzzle element include Crazy Cube [2], the Virtual Kanji Puzzle Game [10], and the Cubed Electronic Puzzle Game from ThinkGeek.com [4, 13].

Crazy Cube is an on-line game where players are asked to "[u]se [their] mouse pointer to link every pair of like coloured markers on the cube faces" [2].

The Virtual Kanji Puzzle Game is an augmented reality application that requires the player to put together basic Kanji parts to form a more complicated one. The idea is to help players learn and practice Kanji.

The Cubed Electronic Puzzle Game, unlike the above two, is a physical puzzle game. It "is composed of four electronic cubes. Each one is surrounded on four sides by magnetic connectors, and each displays 1/4 of the puzzle on its LCD screen." [13]. The objective of the game is to connect the four cubes together so that the combined image will form a meaningful shape or a word.

In summary, there are puzzles that use videos as their image source and there are both virtual and physical puzzles and games that take the form of a cube, but there is no puzzles quite like what is proposed. Also, these puzzles and games are used mainly for entertainment purposes only. Although the moving video jigsaw has an obvious temporal component, no academic studies have been done to leverage that feature.

The following examples are neither puzzles nor games, but they are worth mentioning because they are relevant to the subject.

YouCube is a web application that allows users to map YouTube videos of their choice onto the surfaces of an interactive cube. Besides it not being a puzzle, it is different than the Video Cube Puzzle in that it shows multiple videos concurrently and it uses surfaces, not the entire volume, of the cube to show them [11].

Also not a puzzle, but Cubee [17] and other similar devices, such as pCubee [16] and a view-dependent, polyhedral 3D display [9], are cubic fish tank VR displays. The former two are superior in the sense that they are interactive and can act as input devices, as well as displays. They can be interesting displays and more intuitive input devices than the keyboard and mouse for the VCP. And since both of them use OSG to render their graphics, integration ought to be trivial.

Observations and analyses made from this study can potentially inspire design of a new genre of toys and puzzles [5]. Furthermore, insights gained from this study may also inspire creative applications for new displays such as Cubee and the like.

The remainder of this paper is organized in the following manner: Section 3 describes the Video Cube Puzzle and methodology used to conduct the experiment, Section 4 reports results collected from the user study, Section 5 discusses any unapparent findings and observations from the results, Section 6 draws logical conclusions based on the study results, and, finally, Section 7 concludes this article by suggesting future improvements to the proposed application.

3 Study of Human's Effective Use of Temporal Cues

The research question to be studied is simple: *Can human beings effectively utilize temporal cues provided in the form of video to help them solve a novel 3D puzzle?*

3.1 System

The experiments were conducted running the Video Cube Puzzle on either an IBM ThinkPad with hardware accelerated graphics or on a Windows virtual machine running on an Apple Macintosh. The principle methods of human computer interaction were with a mouse and keyboard, the former is responsible for the manipulation of the individual cubes, while the latter is primarily tasked with program control. **Fig. 3** depicts a participant engaging in a test session.

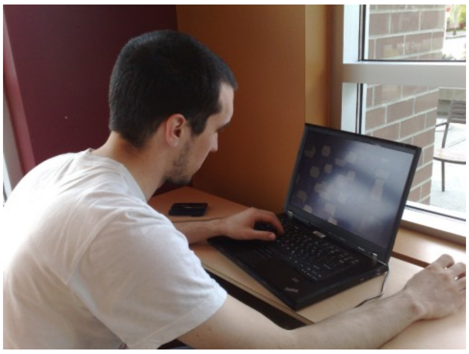

Fig. 3. Participant engaging in a test session.

3.1.1 Design

As mentioned previously, the Video Cube Puzzle is written using OSG. The interface is developed with user-friendliness in mind and it is kept to be as clean and uncluttered as possible. The default interface has two views – a puzzle view, which takes up the majority of the screen real estate, and a much smaller solution view on the bottom-right corner of the screen (Fig. **2**). Each view has its own event handlers and camera manipulator.

The OSG built-in trackball manipulator is adopted for camera manipulation. Also, users can select, rotate, and swap sub-cubes using the mouse. Fig. **4** shows a selected sub-cube with its rotation handles exposed.

In addition to the mouse inputs, there is a set of keyboard commands that apply to the puzzle. Examples include the 'e' key, which toggles the puzzle's expansion feature to expose all faces of the sub-cubes (Fig. **5**), and the 'p' key, which plays the loaded video on the cube (Fig. **6**).

3.2 Experimental Protocol

To mitigate the effects of the "learning curve", much effort has been made to tune the user interactions via keyboard and mouse for the Video Cube Puzzle. For instance, left-clicking on a sub-cube selects it and exposes its rotation handles; right-clicking on the rotation handle rotates the sub-cube in the indicated direction; right-clicking another sub-cube while one is selected swaps the two; left-clicking-and-dragging

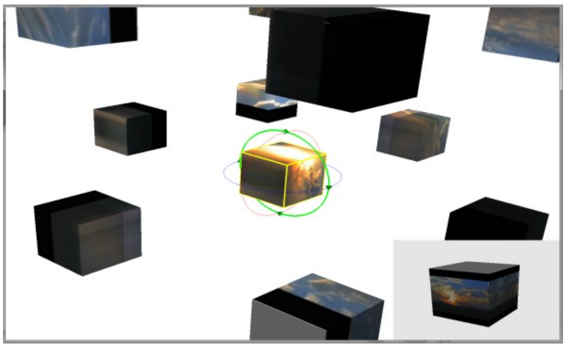

Fig. 4. A screenshot showing a sub-cube being selected, exposing the rotation handles.

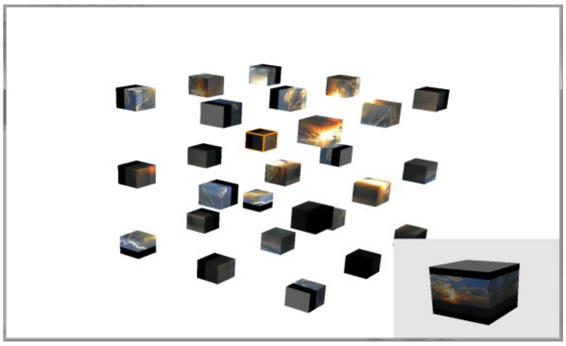

Fig. 5. A screenshot showing the VCP in the expanded state.

Fig. 6. A screenshot of a video being played on the VCP.

rotates the camera around the puzzle; right-clicking-and-dragging pans the camera; and wheel-scrolling zooms the view in and out. Furthermore, tuning of the interface was done at all stages of development so as to continually refine the user experience.

Nevertheless, the user interface requires a certain level of training. As such, before the trials are conducted users are allowed practice "runs" through the system using first a colour only cube, and then with several very trivial cases of the Video Cube, where segmentation variation is held to 2x2x2 and the videos are relatively simple. Users are allowed to repeat such practice runs until they become completely comfortable with the interface.

3.2.1 Baseline

A baseline is established to create a common ground where results collected from users of varying puzzle solving experiences can be compared.

The trial procedure starts with solving a simple 2x2x2 and 3x3x3 cubic puzzles with only solid colours for textures (Fig. **7**). This allows the development of a percent difference for each trial, whereby a comparison can be made with the baseline colour puzzle and any particular video textured puzzle. This is done using the standard percent difference equation applied to a particular trial as follows:

$$\%\,diff = |t_{baseline} - t_{trial}| \,/\, (t_{baseline} + t_{trial}) \cdot 200\ . \qquad \textbf{(1)}$$

This puzzle uses a simple colouring scheme, with six solid colours (i.e. red, blue, orange, yellow, white and cyan) for its six faces, and a solution is reached when all faces are a continuous single colour, as described in [6].

Fig. 7. The Baseline colour cube puzzles against which each of the trials is compared; from left to right are the 2x2x2, 3x3x3, and the completed puzzle.

3.2.2 Surveys

To support the quantitative evaluation outlined in Section 3.2.1, entrance surveys were prepared for each of the candidates in order to assess their previous experience with traditional jigsaw type puzzles, as well as their experience with virtual puzzles. The surveys are designed to determine the user's experience with this type of puzzle, as well as their perception of their own skill at them.

In addition to the entrance surveys, exit surveys are given which have been designed to assess the candidate's feeling with regard to the overall difficulty of the puzzles presented, how they rated the relative increase in the puzzle's difficulty with increased segmentation, and finally how useful the additional temporal element was in solving the puzzle.

3.2.3 Trials

The video were selected to give a broad range of video types from highly regular and predictable to highly random. The goal of selecting these video types was to ascertain the level to which the temporal cues are used as opposed to simply pattern matching the faces of the sub-cubes with the reference cube. Examples of these video puzzles are shown in Fig. **8**.

What follows is a description of the trial procedure used to collect the experimental data.

1. Users were given an entrance survey as outlined in Section 3.2.2 to determine their experience with puzzles.
2. Users were given a learning period as described in Section 3.2 to acclimatize themselves with the user interface, and the puzzle in general.
3. Users were given a 2x2x2 and a 3x3x3 colour-only puzzle to establish a baseline for comparison as described in Section 3.2 and shown in Fig. **7**.
4. Users were given a puzzle which depicts a setting sun on the horizon this video shows constant but highly predictable change in every pixel, this puzzle was given in 2x2x2 and 3x3x3.
5. Users were given a puzzle which depicts a typical film lead in countdown. This video is highly normalized in so far as most areas of the frame remain constant for large periods of time. This puzzle was given in 2x2x2 and 3x3x3.
6. Users were given a puzzle which depicts a rock band playing in a psychedelic forest. This video is highly irregular showing large amounts of variation over time. This puzzle was given in 2x2x2 and 3x3x3.
7. Users were given a puzzle which depicts a murder of crows flying against a grey and highly regular sky. This video creates a non-discernible pattern on all sides of the Video Cube, both spatially and temporally. This puzzle was given in 2x2x2 and 3x3x3.
8. Users were given an exit survey to determine some qualitative measures about their interaction with the Video Cube Puzzle, and the efficacy of temporal information.

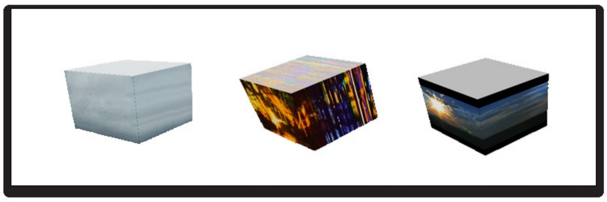

Fig. 8. Examples of the Video Cube Puzzles used.

4 Results

Ten trials were conducted on a broad range of subjects with varying experience with puzzle solving. A chart of results of the average performances is shown in Fig. **9**.

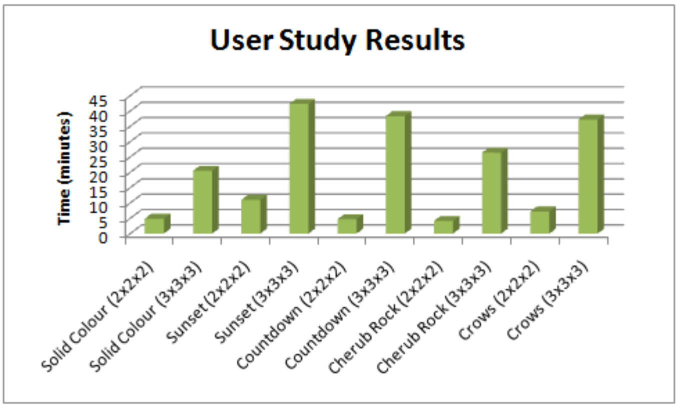

Fig. 9. This chart shows the average times, in minutes, subjects took to complete the puzzles.

4.1 Surveys

Based on our entrance surveys collected from the participants, the average perceived efficacy with standard jigsaw type puzzles was found to be 7.25 (1 being terrible and 10 being very effective), with a standard deviation of 0.71. Furthermore the average proficiency with virtual puzzles (i.e. computer-based puzzles) was 7.625, with a standard deviation of 0.75.

The results of the exit surveys, which pertained specifically to the difficulty of the puzzle, are as follows:

1. Compared to a traditional jigsaw puzzle, on a scale of 1 to 10 rank the relative difficulty of the Video Cube Puzzle (1 being trivial, 10 being extremely difficult). The average response of the candidates was a 7, with a standard deviation of 0.8.
2. If the 2x2x2 Video Cube Puzzle has a difficulty of 1, how would you rate the difficulty of the 3x3x3 (1 being the same, 10 being extremely difficult)? The mean response was 7.4 with a standard deviation of 1.78.
3. How useful did you find the temporal cues given by the video (1 being useless, 10 being critical)? The mean usefulness was seen to be 6.1 with a standard deviation of 2.18.

5 Discussion

While the number of trials conducted seems small, it should be noted that the mean time for completion of the full problem set was 2.9 hours. Furthermore, several candidates found the video puzzle prohibitively difficult to the point of failing to complete the entire problem set. The total mean completion time for a study, finished or unfinished, was 2.42 hours with a standard deviation of 1.06 hours.

Using the formula outlined in Section 3.2.1, relative percent difference calculations were made for each of the trials. A chart of these results is shown in Fig. **10**.

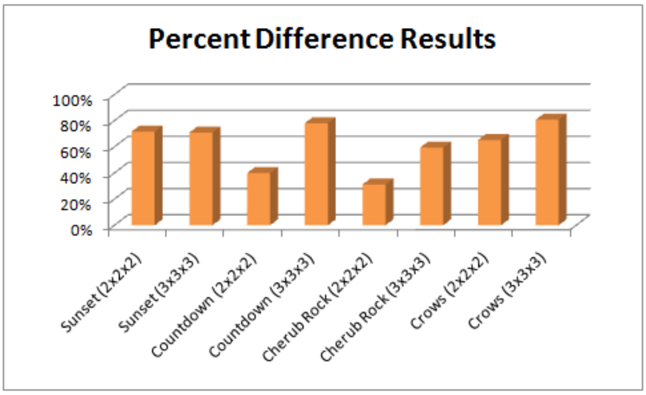

Fig. 10. This chart shows the percent differences of the averages of video cube results and their corresponding averaged baseline solid-colour results.

While the surveyed results show that subjects seemed on average to think the temporal cues were somewhat helpful, observations made by the test givers suggest they seemed more inclined to treat the puzzle as sub-jigsaw puzzles. Furthermore, subjects would often fail to realize the continuum from one sub-cube to another, that is to say two consecutive cubes would have similar textures on their opposing faces. This continuity can be leveraged to more easily solve the puzzle.

The poor performances may also be attributed to an unaccustomed user interface. Two out of ten subjects interviewed expressed frustration with the UI complaining that it did not operate according to conventions established by the CAD tools they were accustomed. Refinement and modification are to be made in a future version to address this issue.

A couple of interesting observations are made from the results: first of all, majority of the participants found the first video-textured cube, namely the Sunset, extremely hard, and then many found the Cherub Rock much easier than the rest. For the Sunset, its results may partly be attributed to it being the first encounter of temporal data and partly to the fact that it is a, both spatially and temporally, slowly varying video; several participants complained about the clouds in the video looking the same across frames. This assertion is further reinforced by the results of the Cherub Rock, which is the video with the most activities and best overall performance. Additional tests using similarly “active” videos will verify this assertion.

6 Conclusions

The results from the experiment have made any conclusive assertions difficult. First and foremost, the fact that 40% of subjects were incapable of completing the puzzle, and those that did take on average 3.05 hours to complete, seems to indicate that, in the general case, randomly selected humans were incapable of leveraging the temporal data provided to assist them in the solving of the puzzle. For a frame of reference, the developers, on average, took 1.66 hours to complete the problem set; however, they had a high degree of familiarity with both the nature of the puzzle and the mechanics of the interface.

The videos chosen were an attempt to ascertain which type of video data would provide the most assistance, and conversely which video data would make the puzzle the most difficult. The Sunset puzzle was seen to be highly temporal, where each pixel in the movie changed in a highly predictable manner, the Cherub Rock video was seen as entirely random thus providing little to no useful temporal data, and the Crows video was seen as containing minimal amounts of temporal data. Based on the results, it appears that the more random and “active” videos provided less of a challenge to the users than the more linear and “predictably” changing videos. Similarly, it was noted that there was a geometric increase in difficulty as the number of segments in the puzzle increased.

Due to the limited sample set, it is at this point not possible to make a definitive assertion about the overall ability of humans to process temporal information when presented in a three dimensional volume. It is the hope of the authors that a future study with a greater sample set could be conducted to obtain more definitive results.

7 Future Work

7.1 Interface

Exit interviews with the candidates resulted in a common feature request – the ability to ``lock" elements of the puzzle once they had been situated. In addition, the controls for rotating individual sub-cubes were said to be too fine, and prone to either mis-clicks, or double-clicks. Refinements on both of these controls and other re-mapping suggested by test participants are required to get more accurate results from users.

7.2 Puzzle Design

The authors have assumed, without grounds, cubic puzzle pieces are appropriate for this study. This fact is to be verified in future studies. Testing using several different geometries, such as pyramids and octagonal prisms, is necessary to gain scientific insights.

Additionally, as discussed in Section 5 testing using more videos, particularly highly varying ones, is required to verify the assertion made about which types of videos would make easier/harder Video Cube Puzzles.

References

1. Audio puzzle, http://www.audiopuzzle.com
2. Crazy cube, http://www.epuzzles.com/crazy-cube/index.htm
3. Jigsaw puzzles free, http://www.jigsaw-puzzles-free.com/VideoJigsaws/Video-Jigsaw-Puzzles.aspx
4. Thinkgeek, http://www.thinkgeek.com
5. Barr, P., Noble, J., Biddle, R.: Video Game Values: Human-Computer Interaction and Games. Interact. Comput., 19(2), pp. 180 – 195. (2007)
6. Berkove, J., Sickle, J., Hummon, B., Kogut, J.: An Analysis of the (colored cubes)^3 puzzle. Discrete Mathematics, 308(7), pp. 1033 – 1045. (2008)
7. Fels, S., Lee, E., Mase, K.: Techniques for Interactive Video Cubism (poster session). In: 8th ACM International Conference on Multimedia, pp. 368 – 370. New York (2000)
8. Fels, S., Mase, K.: Interactive Video Cubism. In: 1999 Workshop on New Paradigms in Information Visualization and Manipulation in conjunction with the 8th ACM International Conference on Information and Knowledge Management, pp. 78 – 82. New York (1999)
9. Harish, P., Narayanan, P. J.: A View-dependent, Polyhedral 3D Display. In: 8th International Conference on Virtual Reality Continuum and its Applications in Industry, pp. 71 – 75. New York (2009)
10. Hirose, J., Guo, Q., Yamamoto, T., Hirayama, M. J.: A Virtual Kanji Puzzle Game Based on 3D Graphics and an Intuitive Inputting Device. In: ACM SIGGRAPH ASIA 2009 Sketches, pp. 1. New York (2009)
11. Meyers, A.: YouCube, http://www.universaloscillation.com/youcube/
12. Parke, F. I.: Adaptation of Scan and Slit-scan Techniques to Computer Animation. SIGGRAPH Comput. Graph., 14(3), pp. 178 – 181. (1980)
13. Piriloo, C.: What's Your Favorite Puzzle Game of All Time?, http://chris.pirillo.com/whats-your-favorite-puzzle-game-of-all-time/
14. Reinhart, M.: Tx-transform, http://www.tx-transform.com/Eng/index.html
15. Sims, L.: Animation, Scan, and Slit Scan Photography. In: 25th Intl. Tech. Comm. Conf. Dallas (1978)
16. Stavness, I., Lam, B., Fels, S.: pCubee: A Perspective-corrected Handheld Cubic Display. In: 28th International Conference on Human Factors in Computing Systems, pp. 1381 – 1390. New York (2010)
17. Stavness, I., Vogt, F., Fels, S.: Cubee: A Cubic 3D Display for Physics-based Interaction. In: ACM SIGGRAPH 2006 Sketches, pp. 165. New York (2006)