Supplementary Material for Videoscapes: Exploring Sparse, Unstructured Video Collections

James Tompkin[1] Kwang In Kim[2] Jan Kautz[1] Christian Theobalt[2]
[1]University College London [2]MPI für Informatik

A Discussion on Refinement of Feature Matching

As discussed in Section 4.1 in the main paper, we have observed that when simultaneously examining more than two pairs of frames, correct feature matches are more consistent with other correct matches than with other incorrect matches. As an example, when frame I_1 correctly matches frame I_2, and frame I_2 correctly matches frame I_3, then it is very likely that I_1 also matches I_3. For incorrect matches, this is less likely. This context information can be exploited to prune incorrect matches.

Similarly to the main paper, results of feature matching can be represented as a graph $\mathcal{G}(\mathcal{F},\mathcal{E})$ which defines frames as its nodes. The existence of an edge between two nodes implies that feature matching for the corresponding frames is valid.

A naive context-based filtering approach would assign a local context-dependent confidence to an edge (I,J) and remove it when the confidence is lower than a threshold. For instance, we could define the confidence of (I,J) as Γ, the degree of overlap of the neighbourhoods $\mathcal{N}_{\mathcal{G}}(I)$ and $\mathcal{N}_{\mathcal{G}}(J)$ of I and J, respectively:

$$\Gamma(I,J) = \frac{|\mathcal{N}_{\mathcal{G}}(I) \cap \mathcal{N}_{\mathcal{G}}(J)|}{|\mathcal{N}_{\mathcal{G}}(I) \cup \mathcal{N}_{\mathcal{G}}(J)|} \tag{1}$$

For instance, if I is neighbouring J and K, it is likely that J and K are each other neighbours (see Figure 1). While this approach may be reasonable when the neighbours of I and J consist of frames in a spatially localized scene, it may mistakenly disconnect I and J if the camera viewpoints are starkly different. For example, a camera operator walks along a path and takes a panning shot from location A, through to location B, and finally to location C. The footage taken from A and B may contain the same landmark. Now consider location C. The footage from B and C overlaps while the footage from A and C does not (see Figure 1). In this case, A and B should not be disconnected just because the subgraph composed of A, B, and C shows low connectivity. The same reasoning continues to cases with more than three nodes.

This specific example can be dealt with by adopting a small threshold value for $\Gamma(I,J)$. However, this may leave incorrect matches in high-density regions. Furthermore, for edges joining nodes in regions with the same density, we could still distinguish correct matches from incorrect ones depending on how these edges are geometrically collocated. In our A, B, C example, the edges joining the nodes are aligned with the same orientation. This orientation consistency and the variations in local density can be used as clues for verifying given connections. To better illustrate this property, let's assume that frames are embedded in a vector space $\mathcal{X}$ which has a metric structure and an underlying probability distribution P. Suppose that distribution P is elongated along a specific axis in $\mathcal{X}$. In this case, an edge parallel to that axis should be more likely to be a correct match than ones oriented orthogonally (Figure 2). In general, the lengths and orientations of edges do not have to be directly related to real geographical locations and camera orientations as in our example in Figure 1.

Given this context, we motivate the use of spectral clustering as follows: given the semi-norm of a vector $f \in \mathbb{R}^n$, whose elements represent the assignment of a cluster index (as a real value, before the quantization by k-means) to each data point:

$$\begin{aligned} \|f\|_L &:= f^\top L f \\ &= \frac{1}{2}\sum_{i,j=1}^{n} k(I_i,I_j)(f^i - f^j)^2. \end{aligned} \tag{2}$$

This norm penalizes the first order variation of f across the set of frames, weighted by k. If we assume that k is (inversely) proportional to a distance in a space embedding, the frames $\mathcal{F}$, $\|\cdot\|_L$ can be understood as a measure of the first order variation weighted by the density of $\mathcal{F}$ in that space. Then, minimizing $\|f\|_L$ tends to place two points I and J in the same cluster (i.e., $|f_I - f_J| \sim 0$) if there is at least one high-density path connecting them (e.g., nodes lying in the upper cluster in Figure 2). Furthermore, when the number of images $n \to \infty$, L converges to the Laplace-Beltrami operator on a compact manifold M in which the data resides [von Luxburg 2007], which is the generator of the diffusion on M. The previously mentioned orientation consistency can be understood in the context of diffusion flow. The corresponding smallest eigenvectors span a subspace of vectors which represent the least penalization by $\|\cdot\|_L$.

In general, the function k is not positive definite (pd) and does not lead to a distance measure. However, the elements of the matrix K are positive and, empirically, the corresponding diagonal elements mostly dominate (i.e., $\sum_i K_{(i,i)} \geq 2\sum_{i\neq j} K_{(i,j)}$). Accordingly, all K in our experiments were pd. When this is not the case, we could instead take its exponential $e^{\beta K} = \lim_{n\to\infty}(I + \frac{\beta K}{n})^n$ with a positive constant β, which is always pd (see [Kondor and Lafferty 2002] for details).

The results of spectral clustering (i.e., the clusters) cannot be used directly to identify portals or supporting sets of frames matching portals. By design, a cluster identified by spectral clustering contains spatially distinct data points. This is not desirable for identifying portals or for identifying sets of appropriate frames to use for the corresponding portal geometry reconstruction. In our A, B, C example, the frames of scene A might not be necessary for the reconstruction of scene C.

We investigated the performance of the graph Laplacian-based connectivity analysis method by comparing it to the local analysis approach. For this algorithm, we randomly sampled 100,000 edges, measured their scores, and removed them when the scores were smaller than a threshold. The threshold was set at 0.4 to result in recalls comparable with our proposed method. Since the order of visiting edges can affect the results, we performed the same experiment 20 times and averaged the error rates. Table 1 shows the results. The precision of local analysis shows an improvement over the results obtained without any connectivity analysis. However, this is still a lower precision and recall than that of our Laplacian-based graph method.

B Time Synchronization

To provide temporal navigation, we perform frame-exact time synchronization between videos in the collection. We group video candidates by timestamp and GPS data if available, and then try to syn-

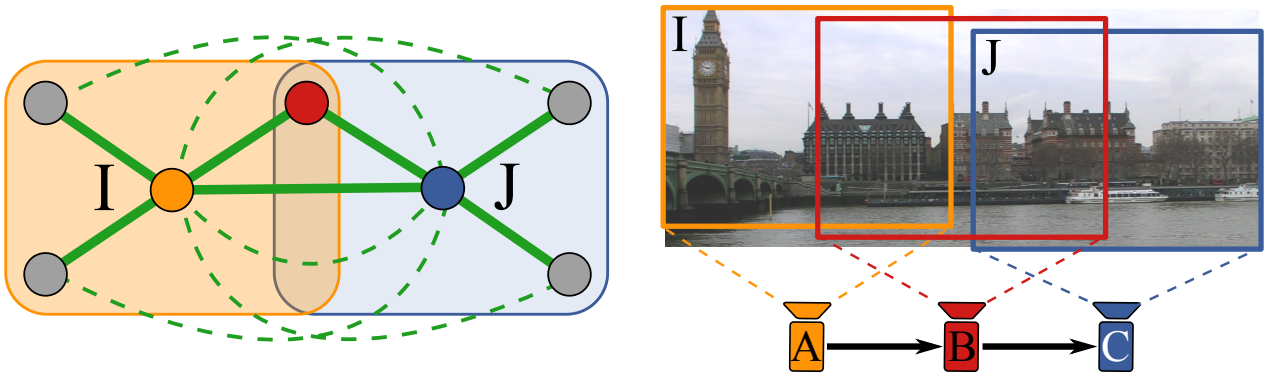


Figure 1: *An example of a hypothetical local connectivity-based confidence assignment (not used in the current system). The diagram on the left shows the subgraph of $\mathcal{G}$ consisting of neighbours of I and J respectively. Solid lines correspond to existing edges, while dashed lines show missing edges which would have supported the edge (I,J). The corresponding confidence value is $\frac{2}{6}$. The diagram on the right shows a case where this confidence assignment would not be applicable (see text for details).*

Phase	Recall	Precision
Spectral analysis	0.53	0.98
Local analysis	0.51	0.95
No connectivity analysis	0.58	0.92

Table 1: *Performance of spectral analysis and local connectivity analysis. 'No connectivity analysis' corresponds to just the holistic and feature matching phases (see Table 2, main paper).*

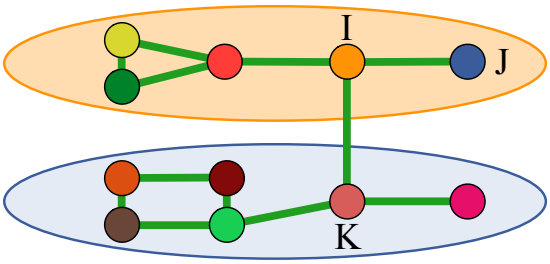


Figure 2: *Example graphical embedding of frames and their connections. Even though (I,J) and (I,K) show the same local connectivity, (I,J) is more likely to be a correct match than (I,K) since the former is in accordance with the flow direction (elongatedness) of the distribution while the latter is not. The underlying distributions P (displayed as ellipses) are not known and should be estimated from frames.*

chronize their audio tracks similar to Kennedy et al. [Kennedy and Naaman 2009]. Videos which are positively matched by their audio tracks are aligned accurately to a global clock (defined from one video at random); hence, portals between these videos create spatial transitions where time does not change (similar to those from Ballan et al. [Ballan et al. 2010]). Videos which are not matched by their audio tracks can only be aligned loosely from their timestamps, and hence create spatio-temporal transitions. This information allows the user interface to optionally enforce temporal coherence among generated tours and to indicate spatial-only and spatio-temporal transition possibilities (Section 5, main paper).

C Discussion of Transitions

C.1 Camera Tracking

In our experiments, KLT feature tracking worked well for tracking videos that are mostly steady. Aligning the KLT features to feature points used in the 3D reconstruction yields smooth sequences of cameras from different videos that are aligned to the 3D geometry with sub-pixel errors. However, in the case of shaky video segments (possibly with rolling shutter artefacts), the quality deteriorates considerably and videos may no longer be accurately aligned with the 3D geometry, leading to ghosting artefacts in the 3D transitions; see videos for Scene 4 in the supplemental material.

For our databases, standard KLT tracking was sufficient for tracking around portals, but other databases may require exposure-compensated KLT tracking. This is a simple swap and does not change any of the computation steps.

C.2 Transition Camera Motion

Ideally, the motion of the virtual camera during the 3D reconstruction transitions should match the real camera motion shortly before and after the portal frames of the start and destination videos of the transition, and should mimic the camera motion style (e.g., shaky motion). To this end, we use the camera poses of each registered video and interpolate them across the transition. This produces convincing motion blending between different motion styles.

C.3 Transition Timing Differences

When constructing all transition types, it is difficult to match their frame timings exactly as they use different techniques to generate new frames. Given a pair of portal frames that are known to visually match, we must choose where in the video transition to place the portal frames. Video frames leading up to and following on from the portal frames may not provide visual matches due to camera motion (such as panning), and so only the portal frames are reliable.

In the simplest case, the cut transition switches at the portal frames. Next, the *dissolve*, *plane*, *ambient point clouds*, and *full 3D – dynamic* transition types all have the pair of portal frames in the middle of their transitions. As the *full 3D – static* transition does not play video through the transition, this type places the portal frames at the start and end of the transition. In this transition type, all geometry through the camera sweep is projected with only the portal

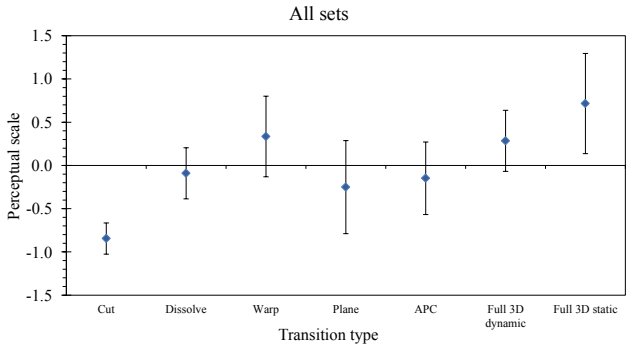


Figure 3: *Perceptual scaling analysis for all scenes/view conditions in our user study.*

Significance	Cut	Dissolve	Warp	Plane	APC	Full 3D dyn.	Full 3D sta.
Cut		2.65E-05	2.89E-05	1.34E-02	3.38E-04	7.17E-06	5.84E-05
Dissolve	2.65E-05		5.57E-02	4.31E-01	5.61E-01	6.49E-02	9.89E-03
Warp	2.89E-05	5.57E-02		2.26E-02	8.20E-02	8.11E-01	1.82E-01
Plane	1.34E-02	4.31E-01	2.26E-02		6.79E-01	7.42E-02	7.46E-03
APC	3.38E-04	5.61E-01	8.20E-02	6.79E-01		3.16E-02	1.23E-02
Full 3D dyn.	7.17E-06	6.49E-02	8.11E-01	7.42E-02	3.16E-02		2.51E-02
Full 3D sta.	5.84E-05	9.89E-03	1.82E-01	7.46E-03	1.23E-02	2.51E-02	

Table 2: *Student's t-test matrix for significance of preference, with $p-value < 0.05$. Green cells denote significantly better, and red cells denote significantly worse. The table should be read as follows: Column 'Cut' with row 'APC' is red, which denotes that Cut is significantly less preferred than APC. Column 'APC' with row 'Cut' is green, which denotes that APC is significantly more preferred than Cut.*

frames. Finally, as the *warp* transition type is image based it starts and ends with the portal frames.

These effects can be seen in the side-by-side comparison of transition types for Scene 3 in our video. In our experiment, we choose to trigger the transitions simultaneously to make it easier for the participants to compare. This leads to some transition clips starting and ending on different frames.

D Transition Experiment Analysis

In Figure 3, we present the perceptual scaling analysis of our user study for all scenes together. Figure 4 shows the perceptual scaling for slight view change cases, and Figure 5 shows the perceptual scaling for considerable view change cases. Table 2 presents the significance matrix for the all scenes case, and Tables 3 and 4 present the significance matrix for the slight and considerable view change cases respectively. Even though the warp transition is not significantly preferred against the full 3D transition types, we believe it is a good choice as the default transition for slight view changes because it a) is the only transition significantly preferred against any other transitions in this case, and b) has a very low perceptual scale variance among participants (Figure 4).

In Figure 6, we show the perceptual scaling analysis for each individual scene and view change of our experiment.

E Label Propagation

We augment the browsing experience by providing semantic labels to objects or locations in videos. This feature has been demonstrated in photo exploration applications [Snavely et al. 2006; Kopf et al. 2008], and we adapt it here to the Videoscape. For instance, if given the names of landmarks, we can allow keyword-based indexing and searching. Viewers may also share subjective annotations with other people exploring a Videoscape (e.g., "Great cappuccino in this café").

A Videoscape provides an intuitive, media-based interface to share labels: During the playback of a video, the viewer draws a bounding box to encompass the object of interest and attaches a label to it. Then, corresponding frames $\{I_i\}$ are retrieved by matching feature points contained within the box. As this matching is already performed and stored during Videoscape computation for portal matching, this retrieval reduces to a fast search. For each frame I_i, the minimal bounding box containing all the matching key-points is identified as the location of the label. These inferred labels are further propagated to all the other frames (matching F_i). If more than two bounding boxes are identified for a single label in a frame then we simply construct a superset box. As the quality of individual key-point matches varies, the inferred bounding box may contain only a part of the object of interest. Thus, we show the center of the box as the location when superimposing tags to video frames (as can be seen in our supplementary video).

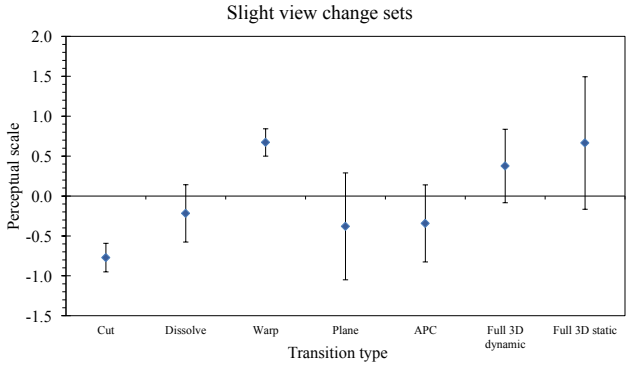


Figure 4: *Perceptual scaling analysis for all scenes with slight view changes in our user study.*

Significance	Cut	Dissolve	Warp	Plane	APC	Full 3D dyn.	Full 3D sta.
Cut		6.58E-03	6.13E-05	2.66E-01	7.16E-02	1.11E-02	3.09E-02
Dissolve	6.58E-03		4.01E-03	6.55E-01	4.84E-01	1.35E-01	1.58E-01
Warp	6.13E-05	4.01E-03		1.08E-02	1.36E-02	3.57E-01	9.86E-01
Plane	2.66E-01	6.55E-01	1.08E-02		9.32E-01	1.92E-01	1.33E-01
APC	7.16E-02	4.84E-01	1.36E-02	9.32E-01		7.06E-02	1.42E-01
Full 3D dyn.	1.11E-02	1.35E-01	3.57E-01	1.92E-01	7.06E-02		3.73E-01
Full 3D sta.	3.09E-02	1.58E-01	9.86E-01	1.33E-01	1.42E-01	3.73E-01	

Table 3: *Slight view change sets student's t-test matrix for significance of preference, with $p-value < 0.05$. Green cells denote significantly better, and red cells denote significantly worse.*

F Study Interfaces

This section presents additional results on video browsing experiments (see Section 6.2 of the main paper). In addition to the question described in Tables 6 and 7 of the main paper, we asked participants 10 questions. The questions were: "For the Videoscapes interface, how useful..."

Q1a: "...were the portal eyes?"

Q2a: "...were grey trails showing from where the video was taken?"

Q3a: "...were the white camera field of views showing from where

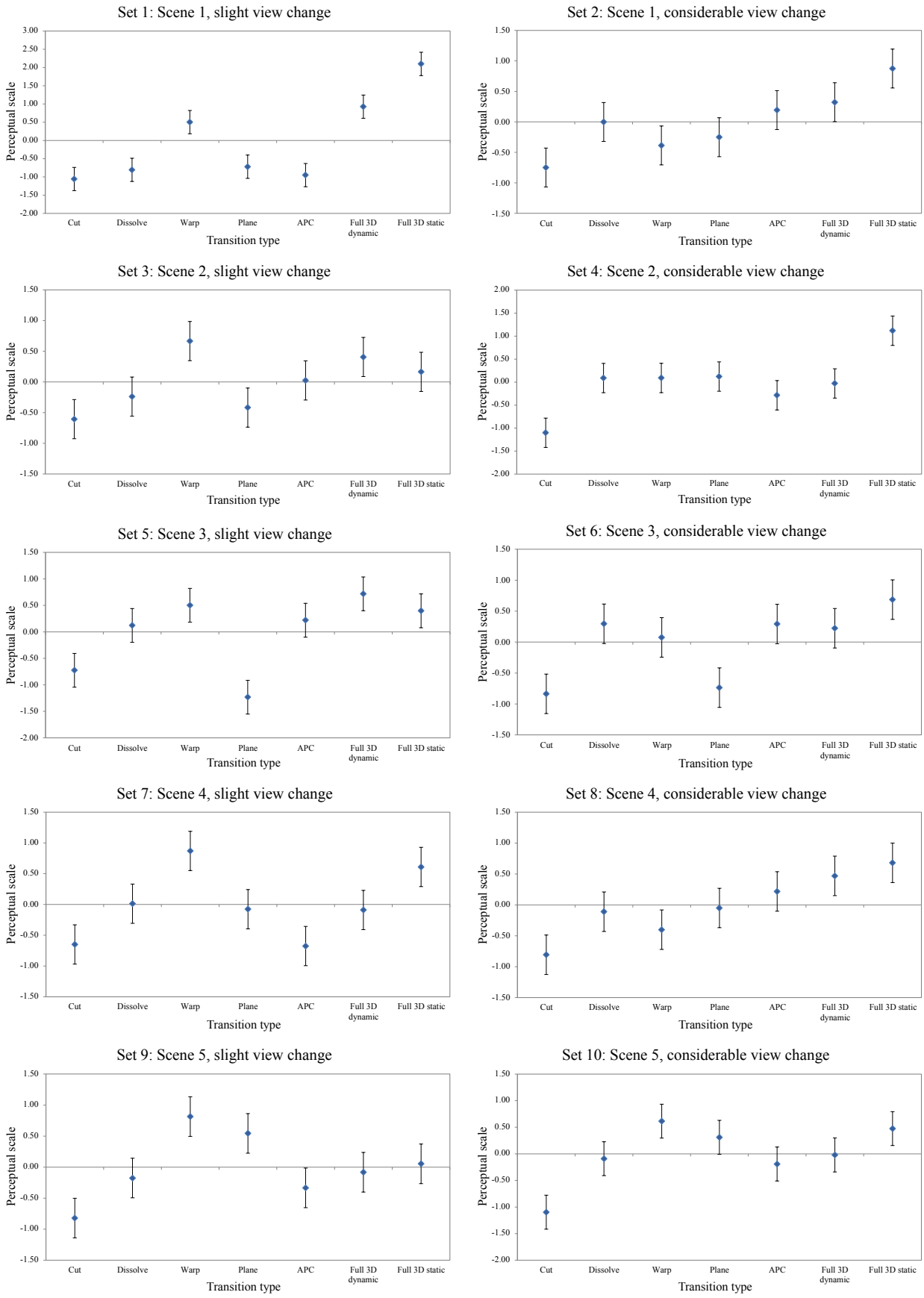


Figure 6: *Scaling analysis for each individual scene and view in our user study.*

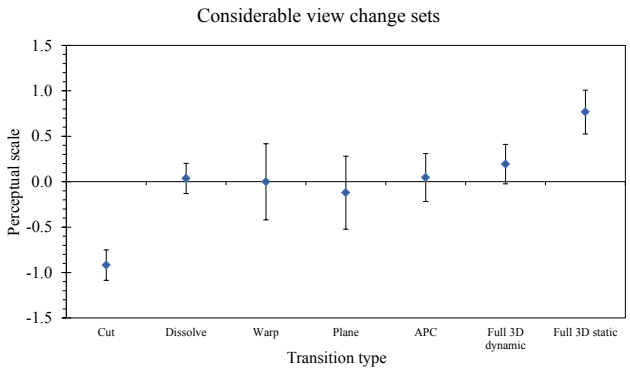


Figure 5: *Perceptual scaling analysis for all scenes with considerable view changes in our user study.*

Significance	Cut	Dissolve	Warp	Plane	APC	Full 3D dyn.	Full 3D sta.
Cut		6.60E-04	2.19E-02	2.84E-02	5.94E-05	1.22E-05	2.34E-04
Dissolve	6.60E-04		8.48E-01	5.46E-01	9.63E-01	3.07E-01	2.94E-03
Warp	2.19E-02	8.48E-01		5.88E-01	8.76E-01	5.18E-01	3.65E-02
Plane	2.84E-02	5.46E-01	5.88E-01		5.90E-01	2.61E-01	1.38E-02
APC	5.94E-05	9.63E-01	8.76E-01	5.90E-01		6.83E-02	1.59E-02
Full 3D dyn.	1.22E-05	3.07E-01	5.18E-01	2.61E-01	6.83E-02		2.06E-02
Full 3D sta.	2.34E-04	2.94E-03	3.65E-02	1.38E-02	1.59E-02	2.06E-02	

Table 4: *Considerable view change sets student's t-test matrix for significance of preference, with $p-value < 0.05$. Green cells denote significantly better, and red cells denote significantly worse.*

the video was taken?"

Q4a: "...was the search box for text searches?"

Q5a: "...was the search box for image searches?"

and "In general, how useful do you think..."

Q1b: "...the portal eyes would be for browsing video collections?"

Q2b: "...the grey trails would be for browsing video collections?"

Q3b: "...the white camera field of views would be for browsing video collections?"

Q4b: "...the search box for text searches would be for browsing video collections?"

Q5b: "...the search box for image searches would be for browsing video collections?"

Table 5 summarizes the results, which support the observations in the main paper that are based on the video browsing experiment: The participants found that interfaces provided by Videoscapes are 'useful'. The portal eyes and image-based searches were especially preferred while the text search and showing camera field of views features were not as preferred as the other interfaces. However, most participants regarded them as useful features for general video browsing systems. This suggests that our prototype interface has some merit beyond the specific task of the experiment.

F.1 Web-based Experiment Components

We show screenshots of the web-based interfaces and questionnaires used by participants in all four of our studies in Figures 7, 8, 9, 10, 11, and 12. The video browsing experiment also included time with each of the three different interfaces tested, and examples of these can be seen in the main paper and supplementary video.

G Photography credits

The photographs of people using video cameras in Figure 1 of the main paper are credited to the following Flickr users:

Man crouching: 'cogdog', `http://www.flickr.com/photos/cogdog/4728847341/`.

Woman in red jacket: 'ramoncutanda', `http://www.flickr.com/photos/ramoncutanda/4096132827/`.

Woman with red hair: 'garryknight', `http://www.flickr.com/photos/garryknight/6667784953/`.

Cyclist: 'goincase', `http://www.flickr.com/photos/goincase/3324038130/`.

References

BALLAN, L., BROSTOW, G., PUWEIN, J., AND POLLEFEYS, M. 2010. Unstructured video-based rendering: Interactive exploration of casually captured videos. *ACM Trans. Graph. (Proc. SIGGRAPH) 29*, 3, 87:1–87:11.

KENNEDY, L., AND NAAMAN, M. 2009. Less talk, more rock: automated organization of community-contributed collections of concert videos. In *Proc. WWW*, 311–320.

KONDOR, R. I., AND LAFFERTY, J. 2002. Diffusion kernels on graphs and other discrete structures. In *Proc. ICML*, 315–322.

KOPF, J., NEUBERT, B., CHEN, B., COHEN, M. F., COHEN-OR, D., DEUSSEN, O., UYTTENDAELE, M., AND LISCHINSKI, D. 2008. Deep photo: Model-based photograph enhancement and viewing. *ACM Trans. Graph. (Proc. SIGGRAPH Asia 2008) 27*, 5, 116:1–116:10.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: exploring photo collections in 3D. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3, 835–846.

VON LUXBURG, U. 2007. A tutorial on spectral clustering. *Statistics and Computing 17*, 4, 395–416.

Video Ranking Experiment

James Tompkin @ UCL, j.tompkin@cs.ucl.ac.uk

This experiment only works reliably in Firefox - if you are having trouble loading videos in another browser, please try with Firefox.
This experiment also requires cookies.

You're using Firefox 6 on Windows!

Scenario:

Imagine you wish to virtually tour a place, such as a city. You are using a new piece of software which can generated a video tour automatically.
You select a path through the city on a map, and with one click the software produces a video tour which moves broadly along your chosen path.
It does this by finding landmarks among a database of videos, and transitioning between the videos at these points.

Here is a short example of the kind of generated tour that you might see:

Instructions:

In the following experiment, you are tasked with ranking a series of videos.
Please rank the videos based on how often you would like to see each kind of transition in your tour.
Placing a video at the top of the list means you would like to see it most often; the bottom of the list is least often.

Please play the video transitions, and change the ranking order by dragging and dropping the videos.
There are 10 sets of videos to rank, and each set contains 7 5-second videos.
Each set covers a different scene, and your judgement should be scene specific.
If you prefer one transition over another for a particular scene, please rank it higher.

Controls:

Use the next and previous buttons to move between sets.
Each video has an associated symbol. These are there to help you remember which video is which.
Please leave any comments for each set of videos in the box at the top of the page. These will be collected automatically.

When you've ranked the last set, press the 'Submit Result!' button to create an email to send me the results.
Your rankings will not be lost as you move between sets.
The experiment will take 30-60 minutes to complete.

Note: there is an loading pause at the start and between sets - your browser has not crashed!

Thank you very much!

Before you begin, please could you describe in a few words your level of expertise with computer graphics and media production:

Start

Clear All Data (Reset Rankings)

Figure 7: *Image of the webpage which explains the experiment to participants. It includes an embedded video showing an example of the transitions that participants are likely to see (in this case, dissolve transitions between video clips that are unused elsewhere in the experiment). We also collect the self-assessed skill level of the participant in media production.*

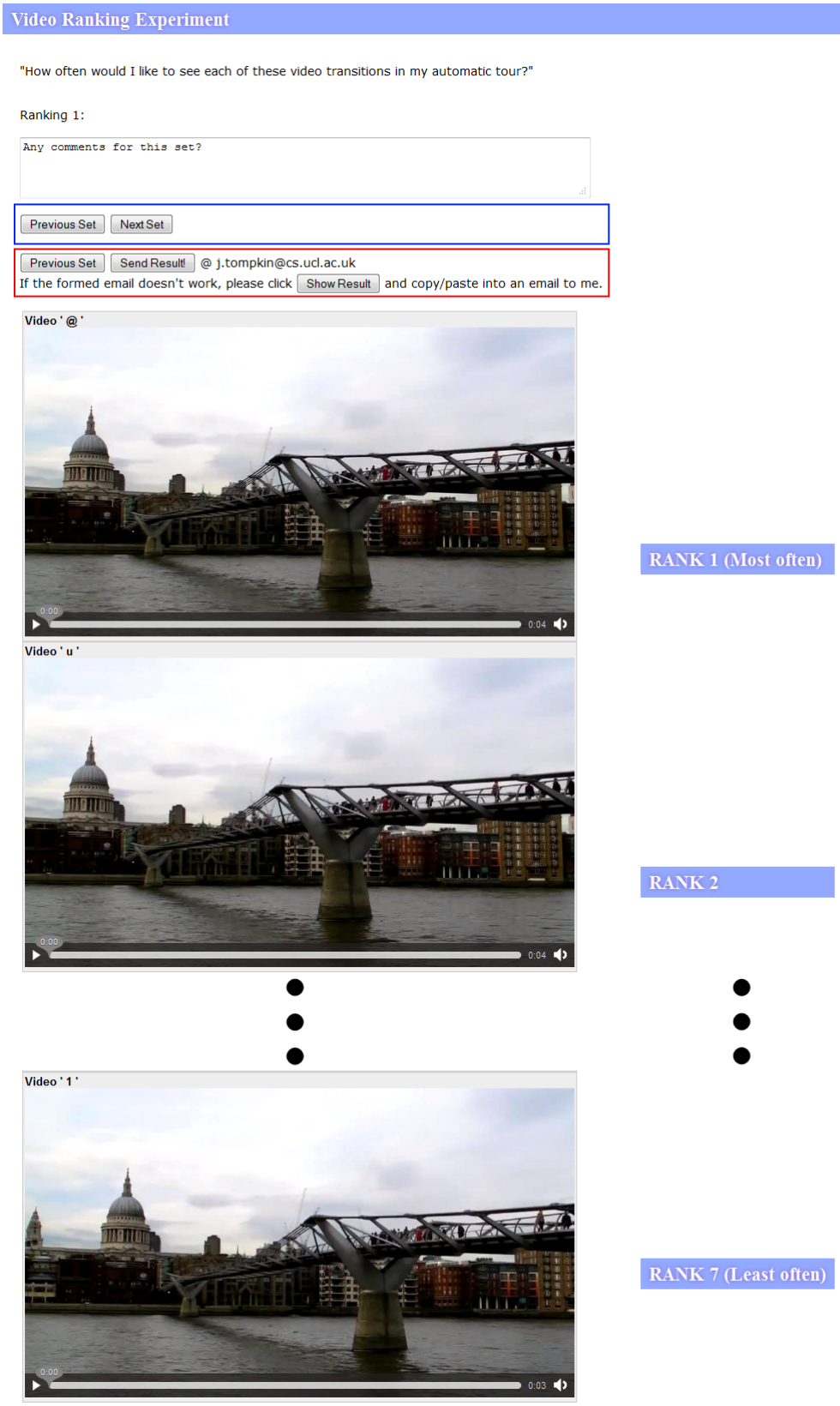
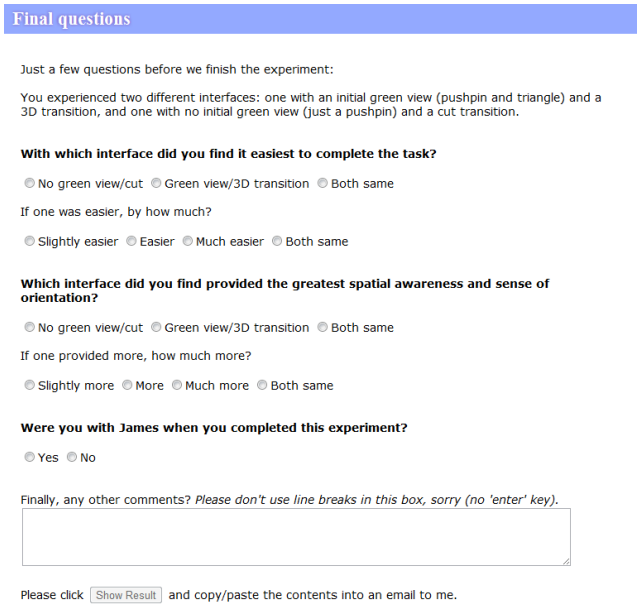


Figure 8: *Image of the webpage for ranking video transitions. Each transition type is randomly ordered onto the page. Participants can drag and drop the videos into order, using the ranking labels to the left to keep track. Comments can be left for each ranking, of which there are 10 in total (5 scenes, each with two view changes) shown in a random order. The region outlined in blue is replaced by the region outlined in red for the final ranking, allowing participants to submit their results remotely.*

Question / Utility	Very useful	Somewhat useful	Not useful	Did not use
Q1a: Portal eyes for task	**14**	0	1	5
Q2a: Grey trails for task	7	4	1	**8**
Q3a: View frusta for task	6	6	1	**7**
Q4a: Text search for task	6	4	0	**10**
Q5a: Image search for task	**11**	1	0	8
Q1b: Portal eyes in general	**13**	7	0	0
Q2b: Grey trails in general	6	**13**	1	0
Q3b: View frusta in general	**10**	**10**	0	0
Q4b: Text search in general	**12**	7	1	0
Q5b: Image search in general	**16**	2	2	0

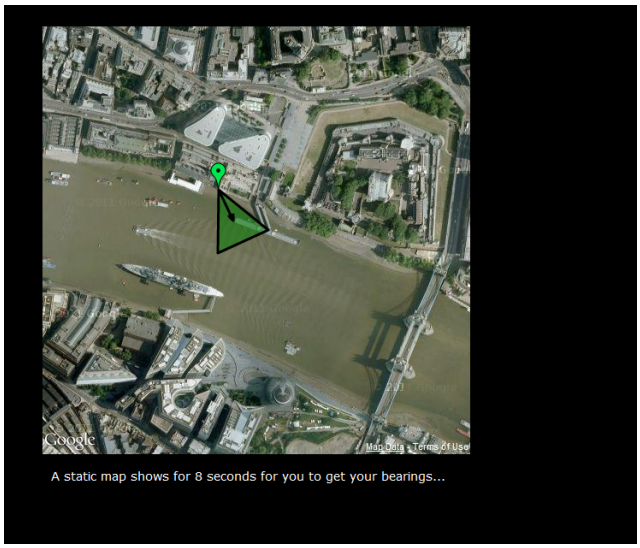
Table 5: *Further questionnaire results of Video browsing experiments showing the number of participants who responded for each choice. Bold signifies the most frequent answer for each question.*

(a) *Initial webpage with explanatory text.*

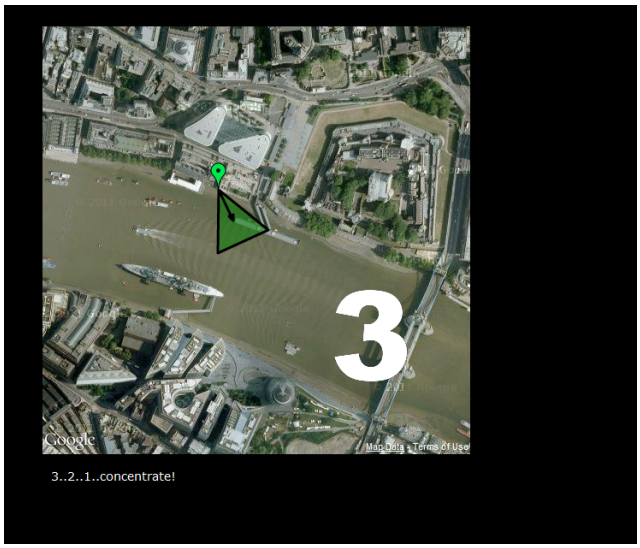


(b) *Questionnaire webpage.*

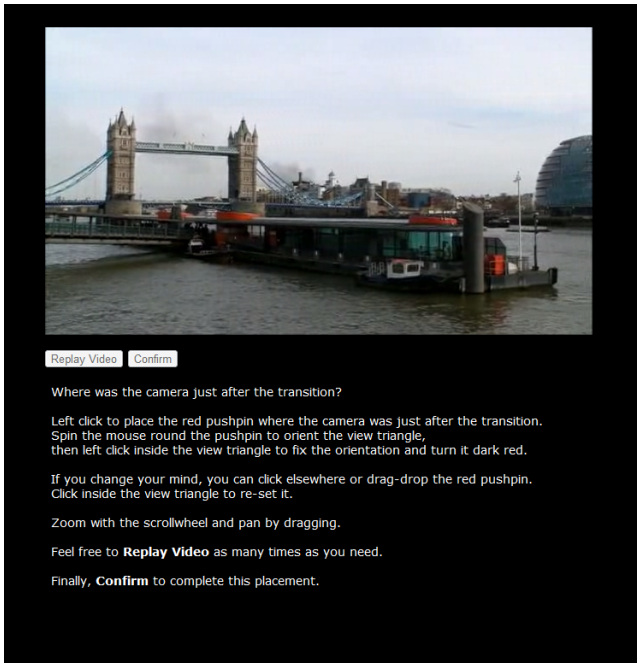
Figure 9: *Spatial awareness experiment website.*



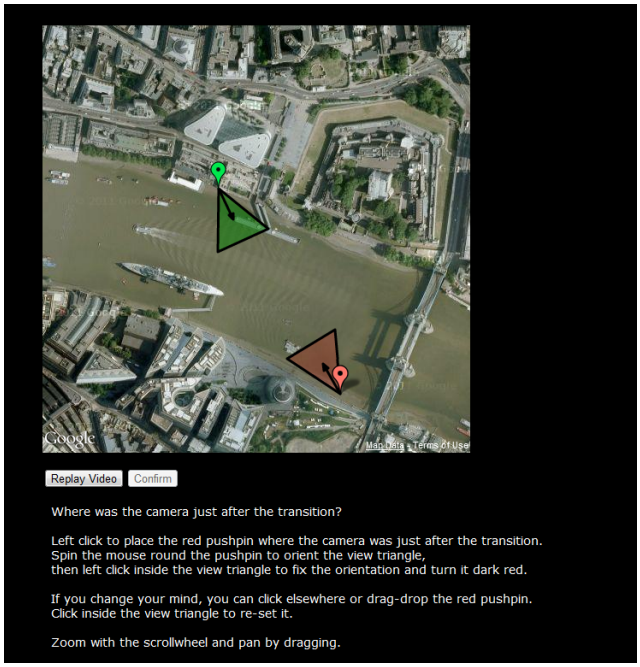
(a) *Participants see a static map, pin, and view frustum for 8 seconds as orientation.*



(b) *A countdown appears for 3 seconds.*

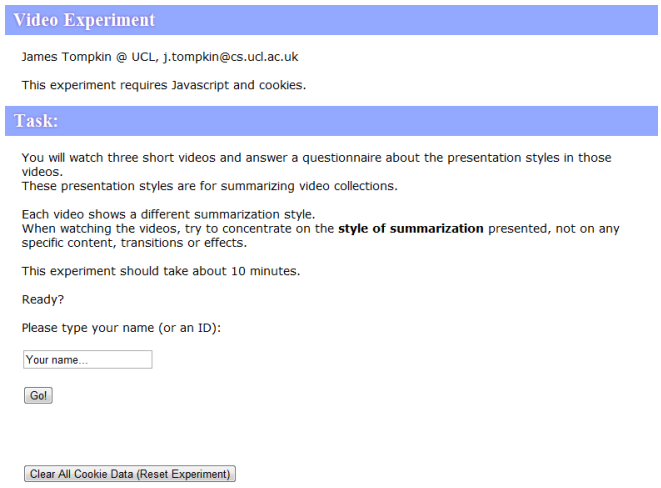


(c) *A video plays, then transitions into another video. This transports the viewer to a new world position and view direction. The two conditions in this experiment show either a cut or a 3D rendered transition.*

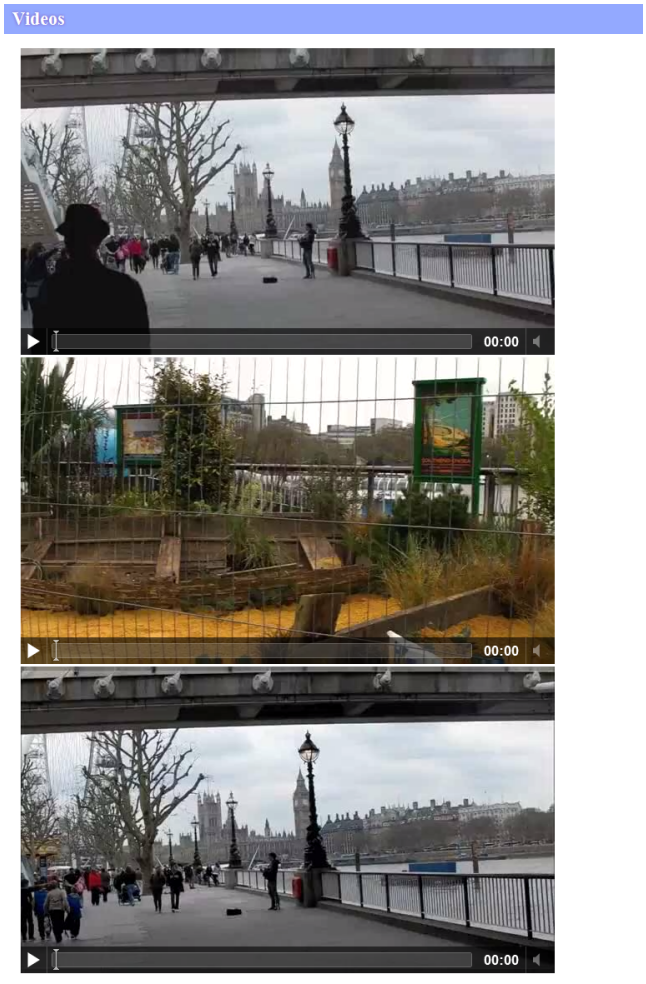


(d) *The participant marks on the map with the red pin/frustum from where they think the second video was taken.*

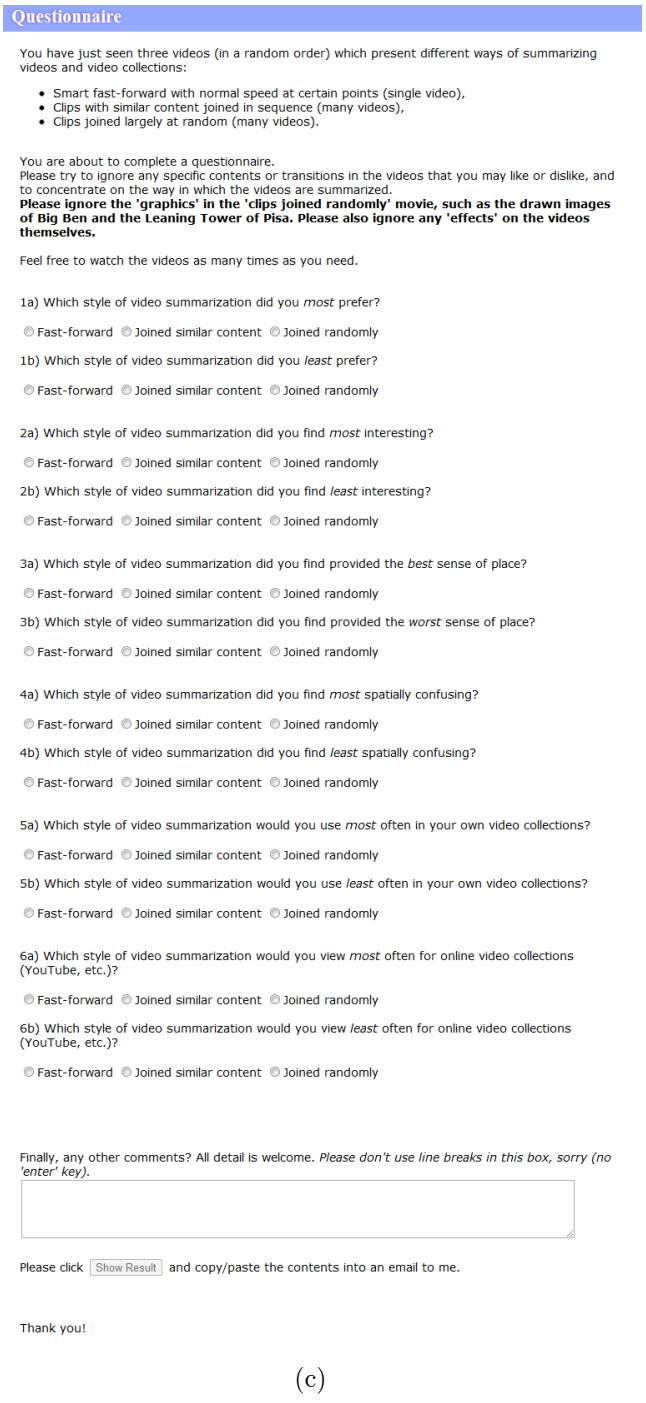
Figure 10: *Spatial awareness experiment website.*



(a)

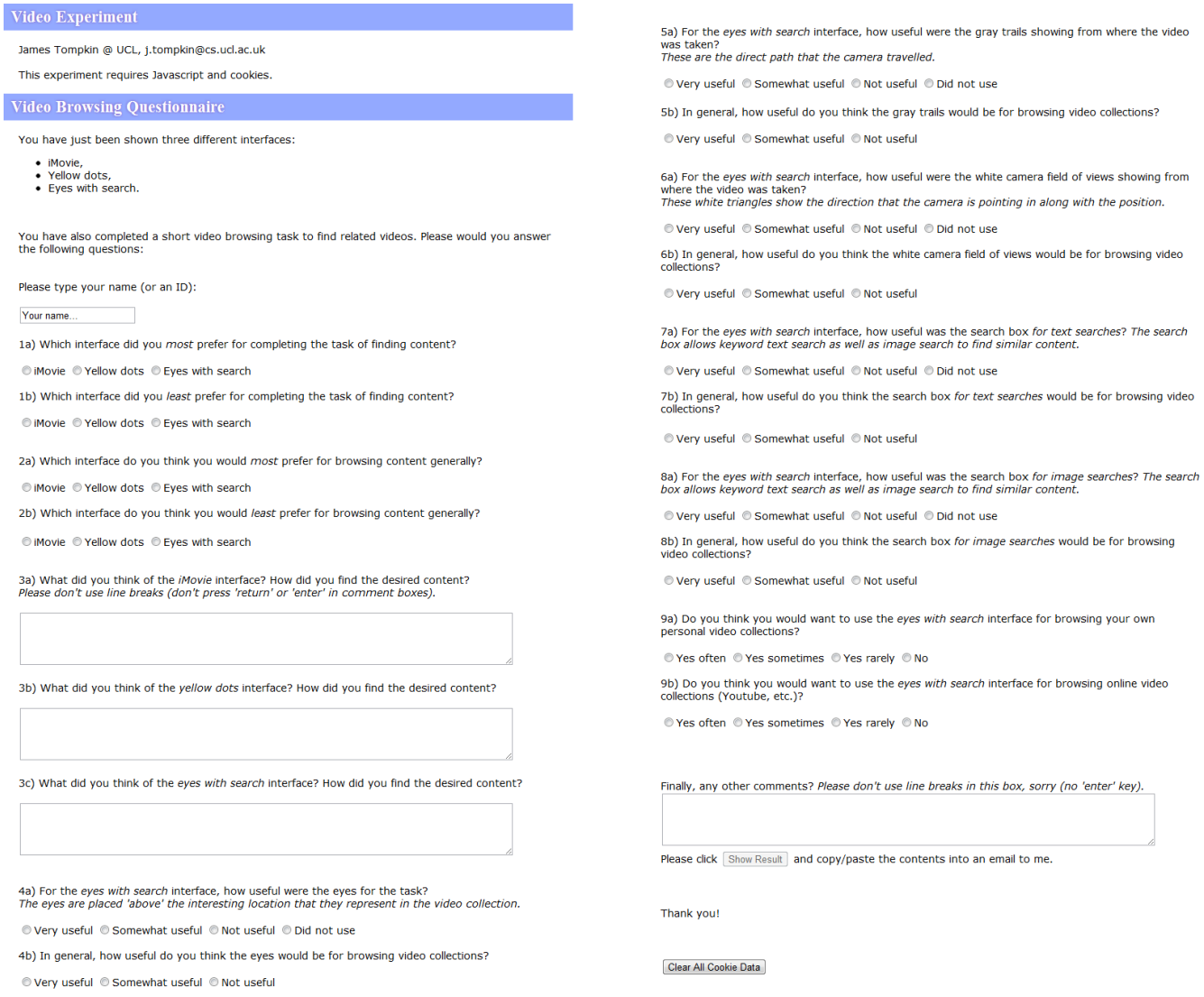


(b)



(c)

Figure 11: *Video tour summarization experiment website.* (a) *Initial webpage with explanatory text.* (b) *Videos which appear in a random order before the questionnaire. These can be replayed at will.* (c) *Questionnaire webpage.*



Video Experiment

James Tompkin @ UCL, j.tompkin@cs.ucl.ac.uk

This experiment requires Javascript and cookies.

Video Browsing Questionnaire

You have just been shown three different interfaces:

- iMovie,
- Yellow dots,
- Eyes with search.

You have also completed a short video browsing task to find related videos. Please would you answer the following questions:

Please type your name (or an ID):

Your name...

1a) Which interface did you *most* prefer for completing the task of finding content?

iMovie Yellow dots Eyes with search

1b) Which interface did you *least* prefer for completing the task of finding content?

iMovie Yellow dots Eyes with search

2a) Which interface do you think you would *most* prefer for browsing content generally?

iMovie Yellow dots Eyes with search

2b) Which interface do you think you would *least* prefer for browsing content generally?

iMovie Yellow dots Eyes with search

3a) What did you think of the *iMovie* interface? How did you find the desired content?
Please don't use line breaks (don't press 'return' or 'enter' in comment boxes).

3b) What did you think of the *yellow dots* interface? How did you find the desired content?

3c) What did you think of the *eyes with search* interface? How did you find the desired content?

4a) For the *eyes with search* interface, how useful were the eyes for the task?
The eyes are placed 'above' the interesting location that they represent in the video collection.

Very useful Somewhat useful Not useful Did not use

4b) In general, how useful do you think the eyes would be for browsing video collections?

Very useful Somewhat useful Not useful

5a) For the *eyes with search* interface, how useful were the gray trails showing from where the video was taken?
These are the direct path that the camera travelled.

Very useful Somewhat useful Not useful Did not use

5b) In general, how useful do you think the gray trails would be for browsing video collections?

Very useful Somewhat useful Not useful

6a) For the *eyes with search* interface, how useful were the white camera field of views showing from where the video was taken?
These white triangles show the direction that the camera is pointing in along with the position.

Very useful Somewhat useful Not useful Did not use

6b) In general, how useful do you think the white camera field of views would be for browsing video collections?

Very useful Somewhat useful Not useful

7a) For the *eyes with search* interface, how useful was the search box *for text searches*? *The search box allows keyword text search as well as image search to find similar content.*

Very useful Somewhat useful Not useful Did not use

7b) In general, how useful do you think the search box *for text searches* would be for browsing video collections?

Very useful Somewhat useful Not useful

8a) For the *eyes with search* interface, how useful was the search box *for image searches*? *The search box allows keyword text search as well as image search to find similar content.*

Very useful Somewhat useful Not useful Did not use

8b) In general, how useful do you think the search box *for image searches* would be for browsing video collections?

Very useful Somewhat useful Not useful

9a) Do you think you would want to use the *eyes with search* interface for browsing your own personal video collections?

Yes often Yes sometimes Yes rarely No

9b) Do you think you would want to use the *eyes with search* interface for browsing online video collections (Youtube, etc.)?

Yes often Yes sometimes Yes rarely No

Finally, any other comments? *Please don't use line breaks in this box, sorry (no 'enter' key).*

Please click Show Result and copy/paste the contents into an email to me.

Thank you!

Clear All Cookie Data

Figure 12: *Questionnaire website for the video browsing experiment. Page appears as one column online.*