# Community Annotation and Remix: a Research Platform and Pilot Deployment

Ryan Shaw, Patrick Schmitz
Yahoo! Research Berkeley
1950 University Avenue, Suite 200
Berkeley, CA 94704-1024
http://research.yahoo.com/location/yahoo_research_berkeley
{rshaw, pschmitz}@yahoo-inc.com

## ABSTRACT

We present a platform for community-supported media annotation and remix, including a pilot deployment with a major film festival. The platform was well received by users as fun and easy to use. An analysis of the resulting data yielded insights into user behavior. Completed remixes exhibited a range of genres, with over a third showing thematic unity and a quarter showing some attempt at narrative. Remixes were often complex, using many short segments taken from various source media. Reuse of spoken and written language in source media, and the use of written language in user-defined overlay text segments proved to be essential for most users. We describe how community remix statistics can be leveraged for media summarization, browsing, and editing support. Further, the platform as a whole provides a solid base for a range of ongoing research into community annotation and remix including analysis of remix syntax, identification of reusable segments, media and segment tagging, structured annotation of media, collaborative media production, and hybrid content-based and community-in-the-loop approaches to understanding media semantics.

## Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems - *Evaluation/methodology*.
H.3.5 [**Information Storage and Retrieval**]: Online Information Services - *Web-based services*.

## General Terms

Human Factors.

**Keywords:** Human-centered multimedia, community media, remix, video annotation, tagging, HCM, UGC.

## 1. INTRODUCTION

The past year has seen an explosion in the amount of video on the web, fueled by the debut of a number of sites for uploading and sharing video clips. As of April 2006, the most popular of these sites was receiving over 35,000 videos per day [1]. At the same time that cheaper bandwidth and disk space have fueled the growth of web video, faster processors and simple editing tools have contributed to the mainstreaming of a "remix culture" in which amateur video editors appropriate and re-create pop culture media. Though these sorts of activities are not new [13], what is new is the scale on which they are occurring.

In parallel to the rapid expansion of media sharing and remix, there is increasing interest in finding ways to harness the collective activity of masses of users in order to produce metadata useful for organizing information resources [11, 28]. At the intersection of these trends lays the possibility of a world in which digital media accumulate layers of explicit and implicit metadata as they are created, shared, and remixed by millions of people [6].

This confluence of media sharing, media reuse, and community annotation raises a number of interesting research issues for human-centered multimedia, including:

- Characterizing community behavior with respect to segmentation, annotation, and reuse.
- Investigating motivation models that promote the creation of high quality annotations and remixes.
- Exploring back-end services that leverage community segmentation and annotation data to provide intelligent tools for annotation, search and retrieval, and remix activities.

To investigate these possibilities further, we have developed a web-based platform that allows users to select, annotate and remix material from a shared media archive. An initial deployment in association with the San Francisco International Film Festival (SFIFF) provided a useful data set for analyzing user behavior, which in turn led to many insights into user behavior and the potential for leveraging community annotation and remix data for a range of purposes.

This paper describes the platform, the pilot deployment, our analysis and ongoing research. The next section places our research in the context of related work. Section 3 presents the architecture of the platform and the implementation of the pilot deployment. Section 4 describes the data collected during the pilot deployment, and details the qualitative and quantitative analyses we conducted on this data. In Section 5, we show how the platform supports ongoing research including investigations into collaborative annotation and authoring, the use of community metadata to simplify media authoring workflow, and the development of new media services and experiences that exploit community-contributed metadata.


*HCM'06*, October 27, 2006, Santa Barbara, CA, USA.

## 2. RELATED WORK

Recent months have seen a host of new web-based video editing applications. Eyespot [9] and Jumpcut [14] allow users to upload home videos and edit them on the web, providing an alternative to simple desktop video editors. Marketers seeking fresh ways to reach web audiences have also jumped on the trend, creating sites that allow visitors to re-edit promotional videos [4]. Although some of these products have a nice UI and rich editing features, the underlying models (especially for annotation and shared community data) are rather simplistic and as commercial platforms do not support research or analysis.

CC-Remix is a web-based system for collaborative remixing of audio [26]. The system is intended for synchronous performance and uses automatic techniques to extract interesting segments for remix, rather than allowing users to produce a collective definition of which segments are interesting.

The talkTV tool exploits closed-captions to allow users to re-edit television content by rearranging lines of dialog, and discusses a number of innovative media services that could be enabled by linking dialog transcripts to video in an open, interoperable fashion [3]. However, the system had minimal support for annotation, and the dependence on textual transcripts limits its usefulness for many kinds of video (e.g., video without dialogue).

These systems all share some of the goals of the platform we have created, in that they emphasize creative, fun exploration of media archives. However, none of them present a thorough analysis of user behavior, and none provide a framework to support research exploring community annotation and remix.

Work towards such a framework is presented in [23], proposing a bottom-up, emergent approach to developing video representation structures by examining retrieval requests and annotations made by a community of video remixers. [18] presents a framework for analyzing and modeling usage of a video archive, but this usage is limited to searching and browsing of video.

## 3. TECHNICAL DESCRIPTION

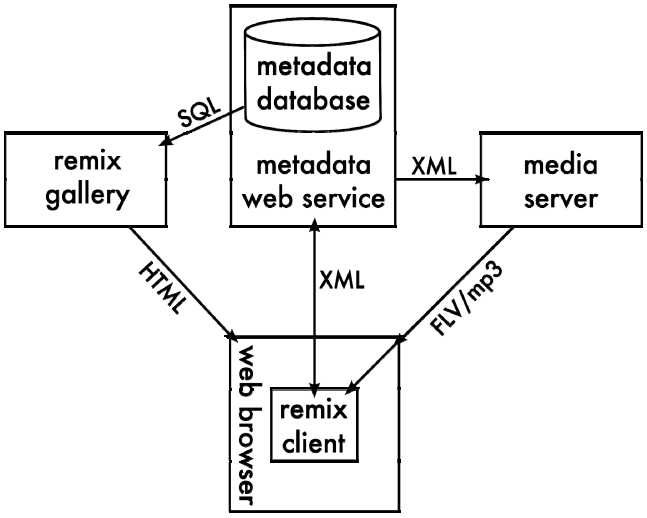


**Figure 1. Components of the International Remix system.**

The International Remix system consists of four components (see Figure 1): a web service and associated database for persistent storage and query of remix metadata, a media server that handles on-demand splicing and serving of remixed media from multiple sources, a client that runs within the end user's web browser, and a gallery site for browsing and viewing submitted remixes. The next four sections detail these components respectively.

### 3.1 Media Metadata Web Service

The web service supports queries for and updates to metadata describing source media objects, media segments, and remixes. It adheres to the REST architectural style for distributed hypermedia systems [10] to ensure scalability and allow the use of standard web intermediaries such as proxy caches.

Following the lead of recent proposed standards for publishing and editing web content [12], the API models media metadata as five *collections* of *resources*: media objects, media segments, remixes, keyword tags, and users (see Figure 2). Each collection and each resource within a collection is identified by a URI. Each individual resource is also identified by a numeric ID that is used to construct a specific URI for that resource.

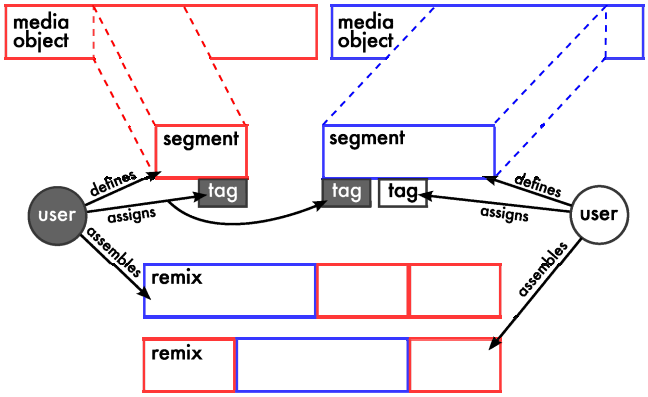


**Figure 2. Resources represented in the metadata web service.**

In general, clients can list resources in a collection, retrieve specific resources from a collection, and add, update or delete a specific resource in a collection. Some collections constrain certain operations, while other collections support additional collection-specific operations on resources.

A typical interaction with the metadata web service begins with an HTTP GET request to the media object collection URI to list the associated media object resources. Such list requests can be filtered and sorted in various collection-specific ways using query parameters appended to the URI. For example, a request for resources in the *media object collection* might be filtered so that only audio objects are listed. The response from the metadata web service is an XML document describing the media objects, including information such as ID, title, author, duration, and the URL at which the media object itself can be found.

Given the ID of a media object, a segment can be created by sending an HTTP POST request to the *media segment collection* URI. The body of the request is an XML document describing the characteristics of the desired segment, including the ID of the media object, start time and end time (for temporal segments), a title, a descriptive note, and a set of keyword tags. The response from the metadata web service is the ID of the newly created media segment. Lists of media segments that have been created for a specific media object can be retrieved via an HTTP GET request to the *media segment collection* URI, specifying the

media object ID in the query parameters. An abstract variant of the *media segment collection* request can retrieve suggested segments based upon community usage or content analysis.

Media segments can be sequenced into remixes by sending an HTTP POST request to the *remix collection* URI. The body of the request is an XML document containing an ordered list of the desired media segment IDs, the (optional) ID of an audio media object that will serve as the remix soundtrack, a title, and a descriptive note. The metadata web service responds with the ID of the newly created remix. The remix can subsequently be updated by sending a new XML description via an HTTP PUT request to the *remix resource's* URI (constructed from the remix ID). Representative poster frames (thumbnail images) for the remix are specified by sending an XML document specifying temporal offsets in the source media and remixed media. The poster frames are modeled as media objects serving as annotations for other media objects.

## 3.2 Media Server

The Media Server delivers media objects, media segments, and remixes over HTTP. To clients it appears to function as a standard HTTP 1.1 compliant web server serving static media files. Unlike a normal web server, however, it can serve segments (specific temporal ranges) of audio and video files as well as remixes (concatenated sequences of segments) from different media files. From the client's perspective, these segments and remixes are indistinguishable from ordinary media files.

The Media Server currently supports Flash Video (FLV) and MP3 audio. For FLV, the server parses the file packets to determine the byte offset at which a packet with a given timestamp occurs. The standard FLV file structure has a header packet followed by a series of stream packets through the end of the file. Each stream packet contains the size of the previous packet, a packet type, the length of the packet and a timestamp in milliseconds.

To stream a portion of a file, the server must patch the timestamps to start at zero for the first packet served. When concatenating multiple segments together it must patch not only the packet timestamps, but also the previous packet size where files have been seamed together. FLV files also contain metadata packets that must be altered to properly represent the stream being served.

When a request for remixed media is received, the Media Server first queries the metadata web service for the segments that constitute the remix. The (HTTP GET) request to the *segment collection* URI specifies the remix ID in the query parameters. The service returns the XML segment descriptions, which the Media Server then uses to splice the appropriate pieces of the source media objects into a remixed media file. This remixed file is then returned to the client that made the original request.

Note that the Media Server's request for metadata is made to the *segment collection* URI rather than the individual *remix resource* URI, because the Media Server needs detailed descriptions of the individual media segments rather than a high-level description of the remix itself. The video server could specify criteria other than inclusion in a specific remix to filter the segment collection, allowing for the creation of dynamic remixes based on segment metadata in addition to explicitly authored remixes. For example, a client might request a remix consisting only of media segments defined by a particular user or tagged with a specific keyword.

## 3.3 Remix Client

The remix client is the primary interface for exploring and segmenting media, and for creating a remix. The initial deployment was developed for use on the San Francisco International Film Festival web site, and focuses on segment definition, remix composition and poster frame selection.

The remix client is a Macromedia Flash application and thus will run within any web browser that supports the Flash Player 8 plug-in. It communicates with the metadata web service using XML messages sent via HTTP and progressively downloads media objects, media segments, and remixes from the Media Server.

When the remix client initializes, it presents the user with a login screen. Upon login, the client queries the metadata web service for lists of source media objects, user-defined segments and in-progress remixes, so the user can continue working where she left off. On the left side of the interface (Figure 3) is a list of source media items represented by thumbnail images. Clicking on a media item loads it into the preview window and displays basic metadata about the item. The video in the preview window can be *scrubbed*[1] by dragging the timeline play head. To select a particular segment of interest, the user drags the triangular *ClipBegin* and *ClipEnd* markers. The frame step buttons allow for fine-grained control over these markers.


**Figure 3. Main interface of the SFIFF remix client.**

Once the user has selected precisely the segment he wants, he can save it to his clip bin by pressing the "Add to My Clips" button, or drag it directly to the remix timeline. This persistently saves the user's segment by posting an XML description of the segment to the metadata web service. A simple drag and drop model provides most of the remix sequencing functionality. Users drag segments onto the remix timeline, and can also re-sequence segments by direct drag and drop manipulation.

In addition to source media segments, the user can define *black segments* (sequences of black frames) and *overlay text segments* (sequences of white text on a black background) and add these to the remix. If musical accompaniment is desired, one of several soundtracks can be selected for the remix. To preclude conflicts

[1] A term from audio/video editing, 'scrubbing' is the playback of media by dragging the timeline position back and forth by hand.

between background and source media audio, the audio tracks of the individual segments in a remix can be toggled on and off.

Pressing the "Play My Remix" button fades the interface into the background and plays the remix that has been created. If the user is happy with what she has created, she can choose to submit her work to a public remix gallery and make it visible to other users. The submission process involves giving the remix a title and description and selecting representative poster frames (Figure 4). Since a remix usually includes media from a number of different sources, a single image would poorly represent remixed content. Our interface allows the user to select up to five poster frames from their remix, and so extends the creative process to the selection of representative images for the remix.

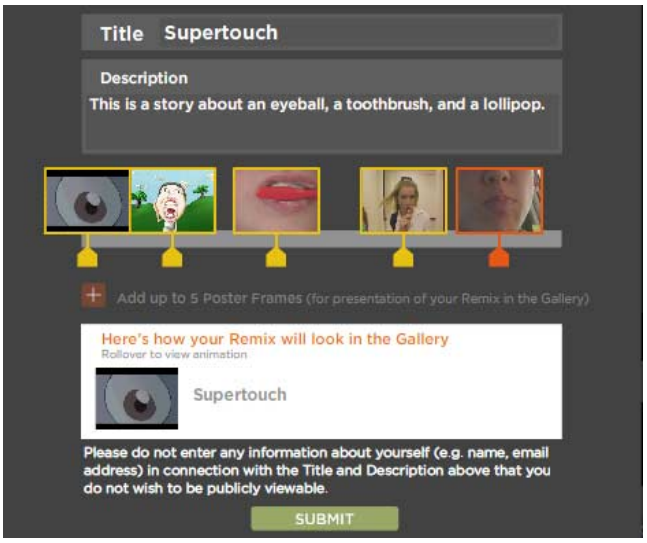


**Figure 4. The poster frame selection interface.**

## 3.4 Remix Gallery

The remix gallery allows visitors to the SFIFF web site to browse and view the remixes submitted for public viewing. Each remix is represented by the poster frames selected by the remix creator. The gallery view cycles through the set of thumbnails when a visitor hovers the mouse pointer over each remix.


**Figure 5. Display of a single remix in the gallery.**

When a visitor clicks on a remix, an overview of the remix is shown (Figure 5). This presents the poster frames (as a spatial montage rather than a temporal one), the remix title, the name of the remixer, and information about source media used including links to more information about the original media. This last item was critical for the SFIFF project so that remixes would drive interest in the films that had contributed source media clips.

# 4. ANALYSIS OF PILOT DEPLOYMENT

In collaboration with the San Francisco Film Society, we conducted a pilot deployment of the remix system as part of the 49th annual San Francisco International Film Festival. By asking directors to permit the creation of remixes from their films and providing tools on their web site to create these remixes, the Film Society hoped to engage younger audiences in the festival and to prompt them to explore festival content in a new way. For us, this was a good opportunity to observe actual usage of the system.

The next sections describe the deployment and present results of qualitative and quantitative analyses performed on the data.

## 4.1 SFIFF Deployment

The remix system was deployed on the festival web site. Nineteen directors from nine countries agreed to contribute short clips from their films. A link from the main festival page led to a gateway page for the remix system, which introduced the remix client and showed a selection of remixes from the gallery. Users could then launch the remix client and try it themselves, or browse the remix gallery to see what others had created.

The remix system was launched three weeks before the start of the festival and ran for five weeks through the festival's end. Midway through, we selected the 50 best remixes and showed them in a special festival screening. The chance to see their work presented in a public forum attracted a number of users to the event, giving us the opportunity to speak with them about their experiences with the system.

Feedback on the system was overwhelmingly positive. Several users commented that they found the system easy to use despite having never edited video before. The compelling content was also mentioned as a positive feature, as it motivated users to spend time playing with various recombinations. Many users also claimed that engaging with the source media clips increased their interest in the films from which the clips were taken.

## 4.2 Qualitative analysis of the remixes

We wanted to analyze the kinds of remixes people create with the tool, and what specific techniques they employed. We also wanted to consider the use of particular features in the tool as part of a usability review. We conducted an initial survey of about 15 randomly selected remixes, and compiled a set of observed qualities and features. Using these, we reviewed all of the submitted remixes and then analyzed the resulting statistics.

We first filtered out remixes with a single untrimmed segment or that otherwise were not meaningful as a "remix" (this *invalid* set was roughly 19% of the submissions, which indicates a need for some UI refinements to preclude such spurious submissions). The remaining *valid* remixes were categorized as either *experimental* or *serious*. We then considered some general qualities of the remix, along with some more specific techniques or effects. Finally, we added criteria to support our usability review.

### *4.2.1 Criteria used for evaluation*

The following are the criteria used, and a brief interpretation of the evaluation intent or criteria. The first group of general qualities included:

- **Attempted narrative**: remix reflects a narrative intent.
- **Thematic unity**: remix segments share a common theme.
- **Funny**: remix reflects a humorous intent.
- **Experiment**: remix appears to be a trivial exploration of the application.

The techniques and effects criteria included:

- **Cutting to music**: remix edit-points closely synchronized to background audio.
- **Clip audio as sound F/X**: segments chosen to produce a sound effect from clip audio.
- **Segment re-use**: some segments used several times (not including the case of looping segments).
- **Segment looping**: some segment repeated as a loop.
- **Uses overlay text**: user-edited text segments in remix.
- **Uses clip language**: leverages segments with text (e.g., signs) or spoken language for narrative/thematic effect.
- **Kuleshov effect**: segment re-use in new context changes meaning/impact of original material [15].

We also considered several criteria related to the user interface:

- **Meaningful remix title**: User found and used the mechanism to set a remix title.
- **Intentional poster frames**: User understood the poster frame mechanism and selected specific frames, rather than just accepting the default first frame of the remix.

### *4.2.2 Results and evaluation*

Tables 1, 2 and 3 present the statistics for each criterion, by group. Note that the values in Table 1 reflect only those remixes that were not considered "experiments" (which amounted to roughly half of the total). Over a third of the serious submissions show thematic unity, roughly one quarter showed at least an attempt at a narrative, but relatively few remixes showed humorous intent.

**Table 1. Remixes by general criteria**

| | |
|---|---|
| Thematic unity | 35% |
| Attempted narrative | 26% |
| Funny | 11% |

Table 2 presents usage statistics for the general editing techniques and effects. These values reflect both the serious remixes as well as the experiments since we were interested in how people approached the general task of remix. We believe that the relatively small number of remixes that were cut to audio reflects the difficulty of this task given the simple tools provided. Nevertheless, several users apparently spent considerable time to produce remixes that were very carefully cut to the background audio. It is also worth noting that nearly half of the submissions used a clip (or clips) several times, and that one third found segments that could be effectively looped.

Many of the remixes leveraged language in one form or another to communicate theme, setting or a particular message. This may reflect the fact that a typical education provides one with a facility for language, but little exposure to film theory or the creative process in other media. For most people, their only experience with creating narrative is either conversational or written – i.e., using language – and so language is an important tool especially for novice visual storytellers.

**Table 2. Remixes by editing techniques and effects**

| | |
|---|---|
| Cutting to music | 7% |
| Clip audio as sound effect | 9% |
| Segment re-use | 45% |
| Segment looping | 33% |
| Uses overlay text segments | 36% |
| Uses clip language | 41% |
| Uses overlay text *or* clip language | 60% |
| Kuleshov effect | 32% |

Not surprisingly, a significant minority of remixes showed clear and sometimes remarkably effective Kuleshov effects. Related to this, a number of users added overlay text segments to convey some specific message, and then chose segments that were related or illustrated this message. Without this overlay text, the intended thematic unity of the resulting remixes would often be unclear, or much less effective.

Table 3 presents the criteria included for usability evaluation. In reviewing the gallery of submitted remixes, we had some initial concerns that certain features (especially setting a project title and choosing poster frames), were too difficult to discover, and so we wanted to gather some statistics to explore this. However, once we broke down the usage by the nature of the remixes produced, we concluded that those users who engaged the tool in a serious manner and completed a proper remix apparently had no trouble discovering and using these features. Moreover, a significant proportion of the non-serious remixes seem to have discovered the poster frame functionality. We think the lower incidence of title usage among the non-serious users may reflect the lack of interest in their final product as much as any usability issue.

**Table 3. Usability criteria seen in remixes**

| | Serious + Experiment | Serious | Experiment | Not a remix |
|---|---|---|---|---|
| Meaningful title | 74% | 94% | 42% | 10% |
| Used posters | 86% | 93% | 76% | 60% |
| Avg # posters* | 2.43 | 2.73 | 1.84 | 1.11 |
| Had bkgd. audio | 78% | 81% | 74% | 52% |
| Had muted audio | 57% | 75% | 30% | 6% |

* when posters used

However, data for several audio-related features seem to indicate that they are not discoverable enough. Users can select a background audio track for their remix, and can then selectively mute the original audio for each video segment in the remix (so that the two audio tracks do not clash). We computed usage of these two features, broken down by the "seriousness" of the remix. As the associated rows in Table 3 illustrate, the background audio feature may have been missed by many of the less serious users, and many fewer of them used the mute feature (as compared to

serious users), indicating a potential problem with that portion of the UI.

## 4.3 Quantitative analysis of the remixes

The metadata database that supports the application also supports analysis of the way users worked with the system to create remixes. The schema was designed to facilitate this analysis (including, e.g., activity time stamps), and so most of the data we needed was readily available with some basic SQL queries. We generally considered several sets within the data: *all* of the remixes created by users vs. the subset that were *submitted* to the gallery (i.e., *completed*); further subsets qualified the submitted remixes as *serious, experimental* or *invalid* (based upon the qualitative evaluation described in Section 4.2, above).

A total of 761 users created 859 remixes, of which 160 were submitted to the gallery. Of these 160, our qualitative evaluation judged roughly 80 to be serious, 50 to be experiments, and 30 to be invalid. We gathered statistics on remix complexity and various aspects of media usage. We also reviewed time spent in the application, and found that most users completed a remix in a single session. About half of the remixes were completed within a half-hour, and 80% were completed within about 2 hours.

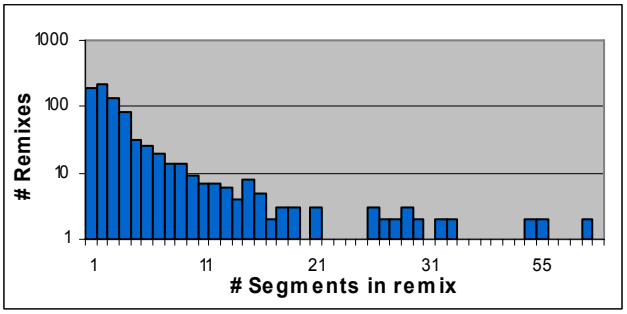


**Figure 6. Segment counts for all remixes created.**

Figure 6 shows the distribution of remix complexity for all remixes (note the logarithmic scale for number of remixes). The average of 4.8 segments per remix, with a median of just 2, reflects the many users that apparently just explored the interface. Over three-fourths of the remixes have 4 or fewer segments.

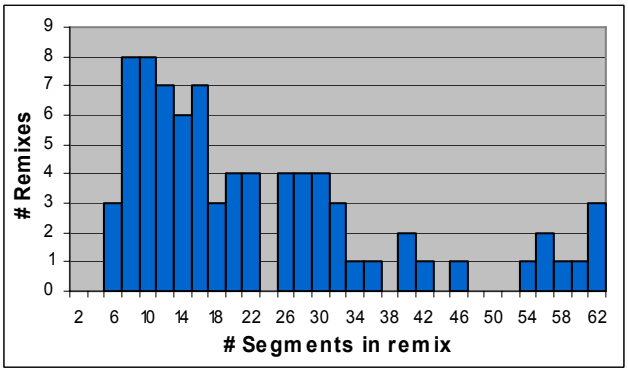


**Figure 7. Segment counts for submitted remixes.**

Figure 7 shows the distribution of remix complexity for 80 *serious* remixes submitted to the gallery, and shows a stark contrast to the collection as a whole. The average segment count for these was 22.3, with a median of 17. The cluster of remixes with segment counts above 50 reflects a particular editing technique – nearly all of these used one or more short segments in a repeating loop, skewing the segment counts. If these looping segments are modeled as a single segment use, the curve looks more regular. Even allowing for this however, the submitted remixes show considerable complexity given the relatively simple tools provided to the users.

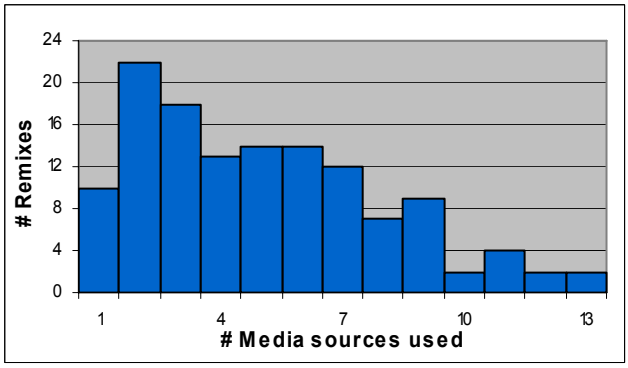


**Figure 8. Media sources used in submitted remixes.**

Figure 8 shows the range of media sources used for *serious* remixes. Many remixes used relatively few sources (the median was 4). This generally reflects the challenge of combining disparate source material. Nevertheless, some users were able to combine a wide variety of sources: half the remixes averaged more than 7 sources, and one in six averaged over 10 sources per remix. There do appear to be limits on this diversity: although there were 19 media sources available, none of the remixes used more than 13.

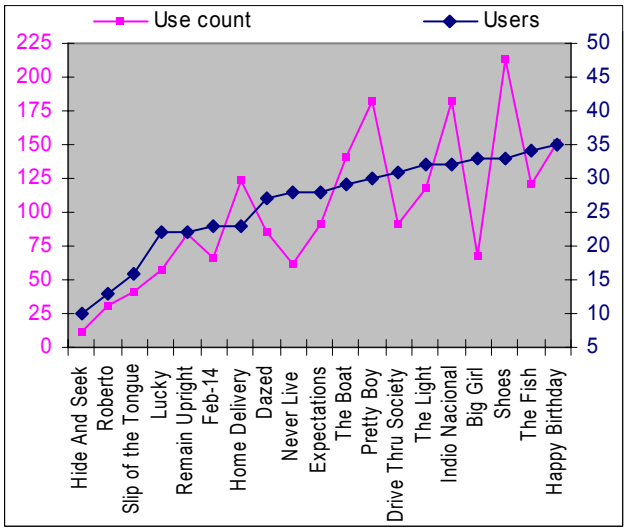


**Figure 9. Media users and segment use counts.**

We also studied the variation in use of media sources, both by the number of unique *users* that remixed some material from each media source, as well as the total number of segments drawn from each source (*use count*). The latter value reflects looping and other multiple use scenarios. These counts were only taken from submitted, valid remixes. Figure 9 presents the results.

The spread of values for *users* is not nearly as broad as the spread for *use counts* (a factor of three vs. an order of magnitude). The two curves are not strongly correlated, and we see a wide variation in the use count per user. The higher values for this ratio seem to correspond to media sources that have a disproportionate use of looping sequences, while lower values may correspond to media for which single, long segments were preferred. We hope to gather more data in future deployments to study this further.

Looking deeper at the way media was used in remixes, we calculated the distribution of segment length for the entire collection of remixes and for the subset of valid submitted remixes. Figure 10 presents the results of this[2]. Although the two graphs are similar in overall shape, several features do emerge. First, the submitted remixes have a higher proportion of shorter segments, reflecting the greater complexity achieved by these authors. The spike around 10 seconds also stands out, and is largely due to the default duration for *black* and *overlay-text* segments – it is not surprising that this is less evident for the more serious remixes. The last distinction relates to the area under the tails of the two curves: 98% of the segments in submitted remixes have durations under about 21 seconds, and 90% of segments are under 10 seconds long. The curve for the full data set shows much longer segment lengths: 98% of the segments are under 38 seconds and 90% are under 17 seconds long.

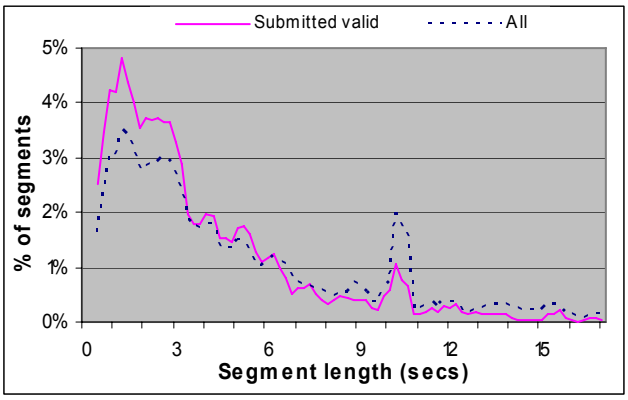


**Figure 10. Segment length distribution.**

Taken together, the media usage statistics show that authors will generally tend to use a limited number of media, and more serious remixers use relatively short segments of media, often sampling many pieces from each media source and using some of these repeatedly.

### *4.3.1 Media reuse histograms*

One advantage of allowing users to select and trim segments by hand is the resulting fine-grained statistics on media usage, unprejudiced by *a priori* shot boundaries. To facilitate analysis of the data, we generated reuse histograms for each source media object. For each 0.1-second interval of the source media object, we counted both the number of different remixes in which the interval was used, and the total number of times the interval was used (to reflect looping and other repetitive usage).

[2] A flat portion of the graph *tail* corresponding to ~10% of the total is omitted to make the majority of the graph clearer.

Qualitative evaluation of the histogram curves yielded a number of interesting patterns. As expected, peaks in the histogram were correlated with points of high emotive energy. The histogram slopes indicate how reuse builds and tapers off as energy builds and wanes. Figure 11 shows a typical example. The source media shot depicts a long zoom toward three flirtatious girls, eventually focusing in on the center girl, who suggestively puts a lollipop in her mouth. The reuse histogram clearly shows an early peak at the point where the three girls flip their hair, and then builds to the main peak at the point where the lollipop is inserted.

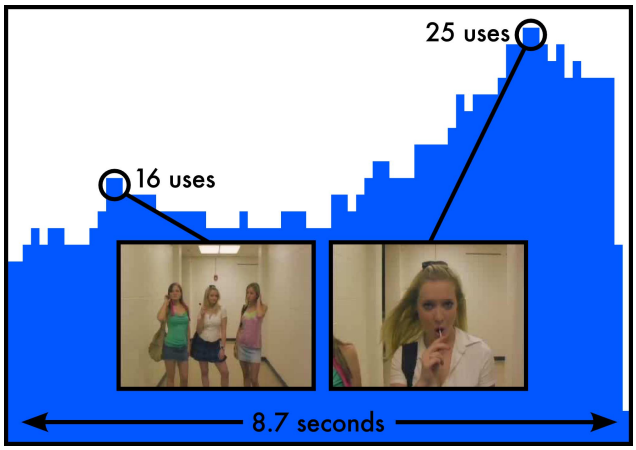


**Figure 11. Reuse building to an emotive peak.**

Figure 12 shows a similar pattern, except that here the emotive energy is concentrated at the beginning of the scene, at the point where the young boy is struck on the head with a stick. We see smaller peaks at important follow-up shots, including the boy's pained reaction, and a look back at the stick-wielding old woman.

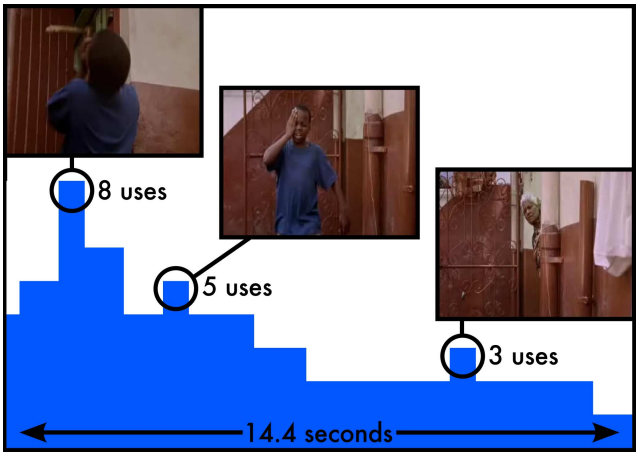


**Figure 12. Reuse tapering off from an emotive peak.**

Contrast these patterns to Figure 13: a low-energy long shot of a man pushing a lawnmower across a field. The wide plateau of the reuse histogram indicates that the shot lacks an emotional focal point; remixers tend to reuse the entire shot and not just a portion of it. Very short source media shots also exhibited plateau-shaped reuse histograms, while longer shots had more exaggerated peaks and valleys. This is likely due to remixers' preference for shorter segments (see Section 4.2.1): longer source shots force a decision about which portion to use, while short shots are just taken as-is.

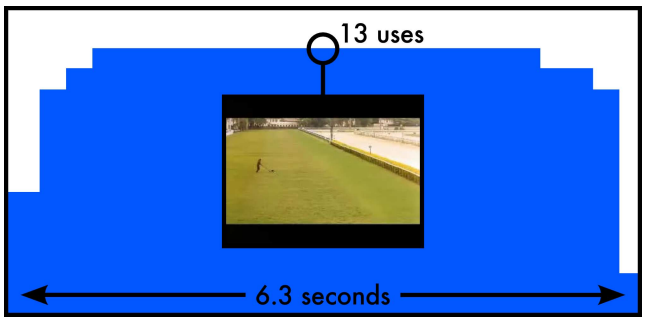


**Figure 13. Consistent reuse across a low-energy long shot.**

As proxies for the emotive impact of media content, these reuse histograms have a number of potential uses for source media summarization and browsing. The amount of reuse can serve as an importance score for selecting representative thumbnails (as is done using content-based techniques in [27]). A scalable video skim [24] could be produced by only including frames above a certain usage threshold, which could then be adjusted for zooming in on scenes of interest.

We also believe that these statistical patterns will have utility for automatically trimming shots. An automatic trimming algorithm that takes community reuse statistics into consideration could trim in such way as to preserve emotive peaks. Such an algorithm would trim the shot in Figure 11 from the beginning, while the scene in Figure 12 would be trimmed from the end. The shot in Figure 13 would not be trimmed at all, reflecting the community consensus that it be used as a whole.

In [20] a method is proposed for obtaining "emotion histograms" to aid retrieval, summarization, and browsing of films, an idea similar to the one presented here. Our method does not classify emotive peaks into specific types. However, our approach gives a finer-grained measure of emotive impact, does not require audio descriptions, and can detect emotionally ambiguous yet clearly powerful points in video content.

Our attempts to characterize these statistical patterns more precisely met with mixed results. Given the clear concentration of source media usage around certain temporal intervals, we were interested in trying to quantify the degree of this concentration and separate the modes of the distribution, in order to develop a collective notion of "interesting" segments. Applying a nonparametric segmentation algorithm [8] to the reuse histograms yielded some tantalizing results, in some cases segmenting the source media into semantically coherent high-level scenes. But in most cases the histograms were not quite peaked enough to allow reliable segmentation of the modes. We also experimented with agglomerative clustering of the segments. This model yielded preliminary results that appear promising for the task of developing a community notion of useful segments. We plan to revisit these lines of investigation in a future study with larger quantities of data.

# 5. BUILDING ON THE PLATFORM

Beyond the collection of data from our initial deployment, our goal in development of this platform was to support continuing research. The platform supports analysis of the syntagmatic aspects of remixes, and of community annotation and tagging support; ongoing work is exploring these areas. The platform can also integrate content analysis tools, supporting hybrid approaches. The next sections describe some of the areas we are actively pursuing.

## 5.1 Analysis of the syntax of remix segments

Analyses of text corpora often focus on common sequences of terms or *n*-grams. *n*-grams can be used as features for clustering or classifying documents, and Markov models can be used to predict the most likely next term given a sequence of terms. Although there have been some attempts to apply *n*-gram techniques to video corpora [21], these have been limited by the fact that visual media, unlike textual media, do not have clearly defined basic units that are used repeatedly within and across documents.

A large corpus of remixed media documents created from a shared archive of media objects and segments, on the other hand, would allow us to identify common sequences of re-used segments. An *n*-gram model trained using such a corpus could potentially be used to implement intelligent authoring assistance to novice remixers. For example, given that the user has assembled a particular sequence of two segments, the model might be used to suggest a commonly used third segment. Alternatively, the model could be used as a sort of "cliché checker" to warn authors away from sequences that are overly common. Such a model could also easily identify segments that are often looped, and automatically offer to repeat these segments for a specified duration.

## 5.2 Identifying reusable segments

Media technology researchers often point to media reuse as a potential benefit of rich metadata. However, even with rich descriptive metadata, identifying media segments that are good candidates for reuse is a difficult problem. An exterior shot of a building may be useful in many contexts for establishing location, but a sign on the building may lower its usefulness (depending on how legible it is). These kinds of distinctions are difficult to capture as continuity constraints or explicit rules for reuse.

Having a corpus of remixed media offers an alternative approach to identifying reusable media segments. Even in our relatively small pilot deployment, certain segments were distinguished by very high amounts of reuse. Of course, a segment may be highly used due to its emotional impact or humor value without necessarily being highly reusable in different contexts. Thus a better metric of reusability may be high incidence of use combined with a high variance in remix contexts.

## 5.3 Adding a tagging interface

Our current metadata schema supports annotation of media items, of temporal segments, and of remixes. Where the initial pilot focused on the segmentation and remix UI, the next prototype user interface adds tagging support. There is an emerging literature exploring different models for tagging and related issues [17], but there is no clear consensus on the best way to facilitate or motivate tag annotations for any given medium or context. We see a number of issues that we would like to explore, both in terms of the user interface as well as the underlying user model. These include:

- Chaptering vs. tagging. Chaptering is the process of dividing a media source into logical pieces, and setting titles for these segments. This is distinct from tagging, and

in some cases may be a pre-requisite. E.g., a UI for audio media requires titles since there is no obvious way to represent the contents of a given audio segment (such as in a list of query results).

- Media vs. segment tagging. In current tagging systems, users typically tag media (e.g., a movie) as a whole, and this raises a number of questions: Will tags on temporal segments within the media work differently than media tags? Will segment tags use the same vocabulary as media tags? Do segment tags reflect a different implicit facet (in a faceted ontology model)?
- Remix tagging. Once users create and share remixes, users may wish to tag the remixes as works. This raises similar questions to the media vs. segment tagging issue.
- Author vs. viewer tagging. In some models, authors do most tagging, while in others, most is done by other users. We are interested in the implications of annotation provenance, and the impact on various aspects of the tagging model.
- Tag suggestion models. Support for tag suggestion can affect the annotations that users create [17]. We are interested both in the development of vocabularies, as well as more general models for motivating annotation activity.

## 5.4 Structured/faceted annotation

Keyword tagging of media objects and media segments lies at the *unstructured* end of the spectrum of user annotation. At the other end of the spectrum is tagging of media using highly structured media description vocabularies. In between these two extremes lies a large design space in which to investigate approaches that attempt to combine the flexibility and ease of use of keyword tagging and the power of structured annotation.

The emergence of actual communities of practice engaged in the appropriation and reuse of media content on a large scale opens the possibility of an empirical approach to the development of these structured representations for media. We are exploring the use of statistical natural language processing techniques to generate and/or extend ontologies from the tag data [22]. In addition to the implicit knowledge capture this supports, we hope to shape the vocabularies and thereby improve the overall metadata collection. We are exploring models in which we iteratively develop these representations in collaboration with communities of users [23]. Eventually we hope to develop tools that will enable the community to participate directly in the ongoing development, maintenance, and evolution of these representations.

## 5.5 Community clip bins

The current remix client only displays segments that the user himself created. However, our metadata schema and API support the creation of remixes from segments defined by other users. This opens the possibility of providing a communal clip bin in which users could potentially see and use all the segments defined by other users. A communal clip bin could ease the task of authoring remixes by allowing users to take advantage of selection and trimming work done by others.

Having a communal clip bin raises a number of interface questions about how to support searching and browsing of segments. Support for segment tagging will be critical for organizing and finding communal segments. User reputation and ranking features familiar from other kinds of social software will also have an important role to play. Tracking and modeling provenance of media segments will also be important once users can directly use and modify one another's assets.

## 5.6 Integration with content-based tools

Implicit and explicit community-generated metadata can help solve many problems that content-based approaches have struggled with. But this does not imply that metadata-based approaches should simply replace content-based ones. On the contrary, there is evidence that the two approaches are complementary. Community-contributed metadata can improve the performance of content analysis algorithms [7], while content-based approaches can in turn improve and enhance tools for annotating and interacting with media [2].

We plan to systematically explore ways in which content-based tools can profitably be integrated with our remix platform. Automatically segmenting source media items at shot boundaries and speaker changes, and automatically tagging high-level events like explosions or lower-level characteristics like camera or object motion vectors could ease the browsing and authoring tasks, allowing us to engage larger communities and thus gather more usage data. However, we must be careful to avoid over-automating the user experience. Robust statistics about the meaning and usefulness of media result from a mass of independent human decisions about what shots to use or how to annotate them. Allowing an algorithm to make these decisions directly conflicts with our goals.

# 6. CONCLUSIONS

Community annotation and remix of multimedia archives are paradigmatic human-centered computing applications, in that their design requires careful attention to user experience and social dynamics. The challenge goes beyond interface and interaction issues to include the question of how to winnow useful patterns from human interactions with media. Addressing this challenge requires real empirical data about what people are doing with media as well as a vision of what people may want to do with media in the future.

An understanding of user motives and behavior will inform the design of improved tools for media annotation and reuse. It can also guide the development of media description standards (e.g., improving playlist formats and incorporating *provenance* of media into remix metadata).

The platform we have developed provides a solid base for ongoing work in human-centered multimedia. The platform provides users with a system for fun, creative exploration of media collections, allowing us to deploy with real communities of users outside of a lab. The platform is instrumented to provide detailed data on usage patterns beyond the basic searching and browsing roles to which users have traditionally been relegated. Finally, we have shown that this data can be used to develop emergent semantics for the media being explored and reused, with implications for new and improved media retrieval, browsing, and authoring applications.

## 7. ACKNOWLEDGMENTS

We would like to thank Sean Uyehara and the San Francisco Film Society for their support, and Joaquin Alvarado and the Institute for Next Generation Internet for hosting International Remix. Thanks especially to Sam Tripodi and the YRB Design team for the pilot design, to Brian Williams, Peter Shafton and our research interns for all their work on the platform codebase, to Jeannie Yang for essential organizational support, and Marc Davis for contributing to our initial conceptualization of the system.

## 8. REFERENCES

[1] Associated Press. Now Starring on the Web: YouTube. *Wired News*, April 9, 2006.

[2] Aurnhammer, M., Hanappe, P., and Steels, L. Integrating collaborative tagging and emergent semantics for image retrieval. In *Proc. of the Collaborative Web Tagging Workshop (WWW '06)* (May 2006).

[3] Blankinship, E., Smith, B., Holtzman, H., and Bender, W. Closed caption, open source. *BT Technology Journal 22*, 4 (Oct. 2004), 151-159.

[4] Bosman, J. Chevy Tries a Write-Your-Own-Ad Approach, and the Potshots Fly. *The New York Times*, April 4, 2006.

[5] Creative Commons, http://creativecommons.org/.

[6] Davis, M. Garage cinema and the future of media technology. *Communications of the ACM*, 40, 2 (Feb. 1997), 42-48.

[7] Davis, M. et al. Towards context-aware face recognition. In *Proc. of 13th annual ACM international conference on Multimedia (MM '05)* (Nov. 2005). ACM Press, New York, NY, 483-486.

[8] Delon, J., Desolneaux, A., Lisani, J.-L., and Patero, A.-B. *A non parametric theory for histogram segmentation.* Technical Report 2005-03, Le Centre de Mathématiques et de Leurs Applications (CMLA), Cachan, France, 2005.

[9] Eyespot, http://eyespot.com/.

[10] Fielding, R. T. Representational state transfer (REST). Chapter 5 in *Architectural Styles and the Design of Network-based Software Architectures*. Ph.D. Thesis, University of California, Irvine, CA, 2000.

[11] Golder, S. A. and Huberman, B. A. Usage patterns of collaborative tagging systems. *Journal of Information Science 32*, 2 (Apr. 2006), 198-208.

[12] Gregorio, J. and de hOra, B. The Atom publishing protocol. IETF Internet-Draft, http://www.ietf.org/internet-drafts/draft-ietf-atompub-protocol-08.txt.

[13] Jenkins, H. *Textual Poachers: Television Fans & Participatory Culture.* Routledge, New York, NY, 1992.

[14] Jumpcut, http://jumpcut.com/.

[15] Kuleshov, L., Kuleshov on Film: Writings by Lev Kuleshov. Translated by Ronald Levaco. Berkeley: University of California Press, 1974.

[16] MPEG-21 Part 5 – Rights Expression Language, http://www.chiariglione.org/mpeg/.

[17] Marlow, C., Naaman, M., Davis, M., and Boyd, D. Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead. In *Proc. of the Collaborative Web Tagging Workshop (WWW '06)* (May 2006).

[18] Mongy, S., Bouali, F., and Djeraba, C. Analyzing user's behavior on a video database. In *Proc. of the 6th international Workshop on Multimedia Data Mining: Mining integrated Media and Complex Data (MDM '05)* (Aug. 2005). ACM Press, New York, NY, 95-100.

[19] Rutledge, L. and Schmitz, P. Improving media fragment integration in emerging Web formats. In *Proc. of the 8th international conference on Multimedia Modeling (MMM '01)* (Nov. 2001).

[20] Salway, A. and Graham, M. Extracting information about emotions in films. In *Proc. of the 11th annual ACM international conference on Multimedia (MM '03)* (Nov. 2003). ACM Press, New York, NY, 299-302.

[21] Satou, T., Kojima, H., Akutsu, A., and Tonomura, Y. Video corpus construction and analysis. *Systems and Computers in Japan 33*, 6 (June 2002), 101-111.

[22] Schmitz, P. Inducing Ontology from Flickr Tags. In *Proc. of the Collaborative Web Tagging Workshop, (WWW '06)* (May 2006).

[23] Shaw, R. and Davis, M. Toward emergent representations for video. In *Proc. of the 13th annual ACM international conference on Multimedia (MM '05)* (Nov. 2005). ACM Press, New York, NY, 431-434.

[24] Sundaram, H., Xie, L., and Chang, S. A utility framework for the automatic generation of audio-visual skims. In *Proc. of the 10th annual ACM international conference on Multimedia (MM '02)* (Dec. 2002). ACM Press, New York, NY, 189-198.

[25] Synchronized Multimedia Integration Language (SMIL) Bulterman, D., et al (eds). W3C Recommendation, 2005.

[26] Tanaka, A., Tokui, N., and Momeni, A. Facilitating collective musical creativity. In *Proc. of the 13th annual ACM international conference on Multimedia (MM '05)* (Nov. 2005). ACM Press, New York, NY, 191-198.

[27] Uchihashi, S., Foote, J., Girgensohn, A., and Boreczky, J. Video Manga: generating semantically meaningful video summaries. In *Proc. of the 7th annual ACM international conference on Multimedia (MM '99)* (Oct./Nov. 1999). ACM Press, New York, NY, 383-392.

[28] Von Ahn, L. and Dabbish, L. Labeling images with a computer game. In *Proc. of the SIGCHI conference on Human factors in computing systems (CHI '04)* (Apr. 2004). ACM Press, New York, NY, 319-326.