MoViMash: Online Mobile Video Mashup

Mukesh Saini
Dept. of Computer Science
National University of Singapore
mksaini@comp.nus.edu.sg

Raghudeep Gadde
Dept. of ECE
National University of Singapore
larst@affiliation.org

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
webmaster@marysville-ohio.com

Lawrence P. Leipuner
Brookhaven Laboratories
Brookhaven National Lab
lleipuner@researchlabs.org

ABSTRACT

This paper provides a sample of a LaTeX document which conforms, somewhat loosely, to the formatting guidelines for ACM SIG Proceedings. It is an *alternate* style which produces a *tighter-looking* paper and was designed in response to concerns expressed, by authors, over page-budgets. It complements the document *Author's (Alternate) Guide to Preparing ACM SIG Proceedings Using LaTeX2$_\epsilon$ and BibTeX*. This source file has been written with the intention of being compiled under LaTeX2$_\epsilon$ and BibTeX.

The developers have tried to include every imaginable sort of “bells and whistles”, such as a subtitle, footnotes on title, subtitle and authors, as well as in the text, and every optional component (e.g. Acknowledgments, Additional Authors, Appendices), not to mention examples of equations, theorems, tables and figures.

To make best use of this sample document, run it through LaTeX and BibTeX, and compare this source code with the printed output produced by the dvi file. A compiled PDF version is available on the web page to help you with the ‘look and feel’.

Categories and Subject Descriptors: I.2.10 [Vision and Scene Understanding]: Video Analysis.

General Terms: Algorithms, Design.

1. INTRODUCTION

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes (for instance, 9 point for body copy), a specified live area (18 × 23.5 cm [7” × 9.25”]) centered on the page, specified size of margins (1.9 cm [0.75”]) top, (2.54 cm [1”]) bottom and (1.9 cm [.75”]) left and right; specified column width (8.45 cm [3.33”]) and gutter size (.83 cm [.33”]).

during the summer of 2011, validating the significant increase in the amount of mobile video uploaded to PhotobucketâĂŹs video shar- ing website (14ÃŮ in Summer 2011 compared to December 2010) [2]. Proliferation of such mobile devices with video capture capa- bility has enabled users to capture video of their life events such as concerts, parades, outdoor performances, etc, and socially share them with friends and public as it happens. Videos recorded by a single user at such events are shot from a limited range of an- gles and distances from the performance stage, as an attendee typi- cally has limited mobility (e.g., constraint by seating arrangement). The recorded video can be monotonous and uninteresting. Further- more, videos recorded are typically short (in the order of minutes or tens of minutes), due to tired arms or power constraint of mobile devices. There are, however, likely to have more than one users recording the same performance from different angles at the same time, especially at a well-attended performance. These recorded and shared video clips of the same performance can be cut and joined together to produce a new mashup video, similar to how a TV director of a live TV show would switch be- tween different cameras to produce the show. Generation of a video mashup can be cast as a video selection problem: given a set of video clips capturing the same performance event, automatically select one of the video clips at any one time instance to be included in the output mashup video. In this paper, we introduce MoViMash, our approach to solve the above video selection problem. MoViMash aims to produce mashup video from a set of mobile devices that is interesting and pleasing to watch, and uses a combinations of content-analysis, state-based transitions, history-based diversity, and learning from human editors to achieve this goal. We now provide an overview of how MoViMash works in the usual setting of live performances, shown in Figure 1. There is generally a staging area and an audience area where the audiences either sit or stand to watch the performance, and record the perfor- mance with a mobile device. This setting poses a few challenges

WOODSTOCK ’97 El Paso, Texas USA

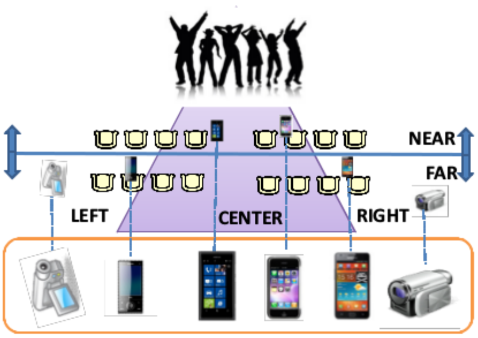


Figure 1: A general performance scenario

The good news is, with only a handful of manual settings, the LATEX document class file handles all of this for you.

The remainder of this document is concerned with showing, in the context of an "actual" document, the LATEX commands specifically available for denoting the structure of a proceedings paper, rather than with giving rigorous descriptions or explanations of such commands.

The shooting angle of the remaining videos are then classified as either center, left, and right; and distance from the stage as near and far as shown in Figure 1. This classification is done every time we perform video selection because mobile users may change their po- sition over time. MoViMash now decides which shooting angle and distance should be used; and for how long the selected class should persist. To this end, MoViMash tries to imitate a professional video editor, by using a finite state machine, whose transition probabilities are learned from analyzing professionally edited videos of the same type of event. The rationale behind the inclusion of learn- ing is that, we have observed that there are no generic editing rules that can be precisely defined to work with all types of events. The video editors make fine decisions such as shot lengths and transi- tions based on their experience which is hard to enumerate. The videos from the selected class are further ranked based on the video quality and diversity values to make the final selection. To consider video quality, MoViMash favors video with low blurri- ness, low blockiness (good compression), good contrast, and good illumination in each video. To consider diversity, MoViMash stores a history of recent video selections and favors videos with dissimi- lar views with recent selections. We have developed MoViMashâĂŹs algorithm such that it is online and only depends on history information. As such, even though it is not our main goal in this paper, MoViMash can be applied to mashup of live video feeds from mobile devices. We now briefly compare MoViMash to existing work to high- and only depends on history information. As such, even though it is not our main goal in this paper, MoViMash can be applied to mashup of live video feeds from mobile devices. We now briefly compare MoViMash to existing work to high-

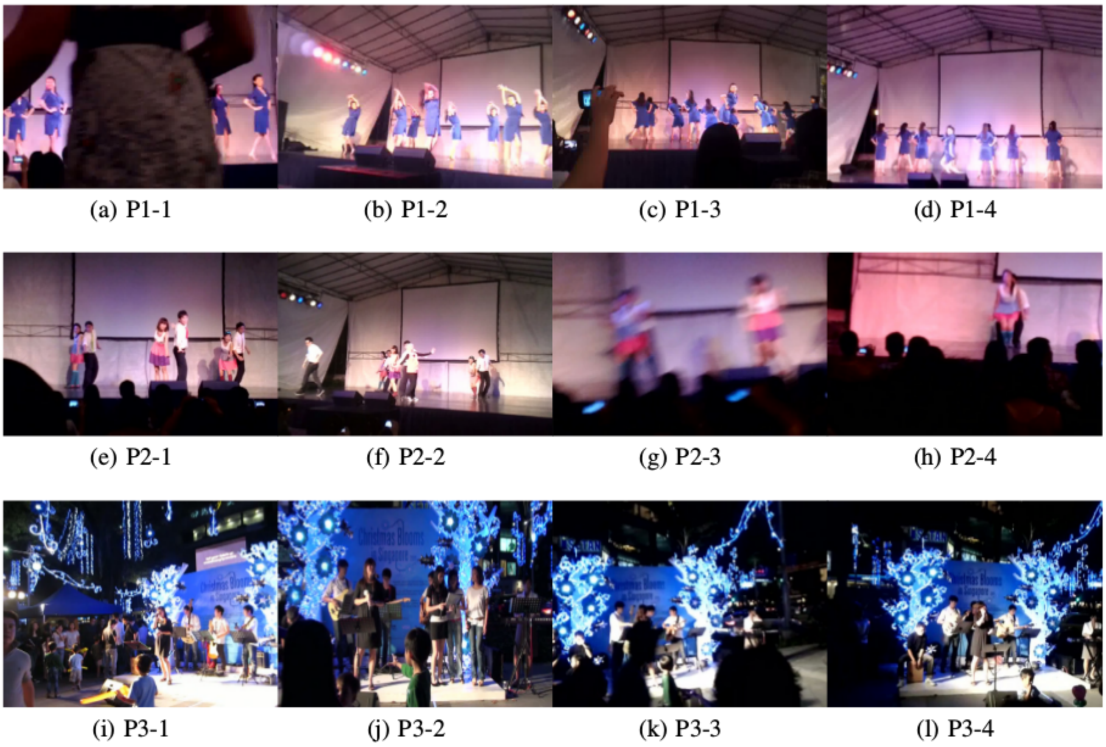
Contributions. We now summarize our contributions in this pa- per as follows:

- We propose a state-based approach for shot selection that
- We propose a state-based approach for shot selection that
- We propose a state-based approach for shot selection that
- We propose a state-based approach for shot selection that

The shooting angle of the remaining videos are then classified as either center, left, and right; and distance from the stage as near and far as shown in Figure 1. This classification is done every time we

2. PREVIOUS WORK

There has been only few works on online camera selection. In most of these works, videos are mainly selected to show the speak- ers. In the work by Machnicki and Rowe [9], an online lecture webcast system is presented in which the cameras that are focusing on speaker and the presentation (the screen) are selected iteratively until anybody from audience asks question. When audience ask question, the camera that is focusing the person asking question is selected. The automatic selection of cameras in a lecture webcast is extended by Zhang et al. [21] to include audio based localization and speaker tracking. Similar approach is taken by Cutler et al. [6] in a meeting scenario where camera that shows the current speaker is selected. Ranjan et al. [12] use face tracking and audio analy- sis to show the close-up of the person talking. Since performers play more important role than speakers in live concerts, a speaker based selection is not appropriate. Further, the faces are generally far from the camera which cannot be detected. Therefore, face de- tection is not a reliable basis to select videos. Al-Hames et al. [3] extends the camera selection work to include the motion features. We do not use motion features in our frame- work because both performers and audience generate continuous motion. Also, the movement of the mobile camera can inject erro- neous motion in the video, which is aesthetically appealing. Yu et al. [20] propose to customize the camera selection and shot lengths based on user preferences. At every lecture webcast receiving site, the user can give score to the videos and specify rules for shot lengths. While such an interactive selection of cameras is useful for educational scenarios, people may find it annoying and stress- ful for performances, particularly when the number of videos is large. A camera selection method for sports video broadcast is proposed by Wang et al. [16]. The authors assume one main cam- era and other sub cameras. The empirical main camera duration is found to be from 4 to 24 seconds, and sub camera duration is found to be 1.5 to 8 seconds. They select a sub camera based on the clar- ity of the view, determined using motion features. In our work, along with shakiness of the videos, we also calculate view qual- ity in terms of occlusion and rotation; and video quality in terms of contrast, blur, illumination, and blockiness. We also include explicit measurement of diversity in the framework. A camera selection method for sports video broadcast is pro- posed by Wang et al. [16]. The authors assume one main cam- era and other sub cameras. The empirical main camera duration is found to be from 4 to 24 seconds, and sub camera duration is found to be 1.5 to 8 seconds. They select a sub camera based on the clar- ity of the view, determined using motion features. In our work, along with shakiness of the videos, we also calculate view qual- ity in terms of occlusion and rotation; and video quality in terms of contrast, blur, illumination, and blockiness. We also include explicit measurement of diversity in the framework. Engstrom et al. [8]



(a) P1-1 (b) P1-2 (c) P1-3 (d) P1-4

(e) P2-1 (f) P2-2 (g) P2-3 (h) P2-4

(i) P3-1 (j) P3-2 (k) P3-3 (l) P3-4

Figure 6: Selected frames from the recordings: (a-d) P1, (e-h) P2, (i-l) P3

heightheight

Table 1: Frequency of Special Characters

Non-English or Math	Frequency	Comments
Ø	1 in 1,000	For Swedish names
π	1 in 5	Common in math
\$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage

discuss automatic camera selection for broadcast in a sports event capture scenario. The work mainly promotes collaborative video production, i.e., video recorded by production team as well as the consumers.

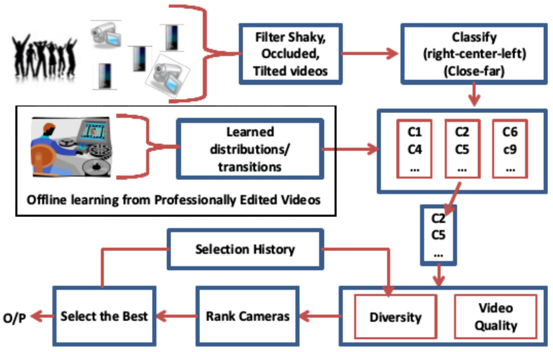

Filter Shaky, Occluded, Tilted videos
Classify (right-center-left) (Close-far)
Learned distributions/ transitions
Offline learning from Professionally Edited Videos
C1 C4 ...
C2 C5 ...
C6 c9 ...
C2 C5 ...
Selection History
O/P
Select the Best
Rank Cameras
Diversity
Video Quality

Figure 2: The virtual director framework

2.1 Type Changes and *Special* Characters

We have already seen several typeface changes in this sample. You can indicate italicized words or phrases in your text with the command `\textit`; emboldening with the command `\textbf` and typewriter-style (for instance, for computer code) with `\texttt`. But remember, you do not have to indicate typestyle changes when such changes are part of the *structural* elements of your article; for instance, the heading of this subsection will be in a sans serif[1] typeface, but that is handled by the document class file. Take care with the use of[2] the curly braces in typeface changes; they mark the beginning and end of the text that is to be in the different typeface.

1. **Filtering:** In the filtering step, we determine videos that are unusable by comparing occlusion, shakiness, and tilt scores against empirically determined thresholds. The remaining videos are passed to the classification stage.

2. **Filtering:** In the filtering step, we determine videos that are unusable by comparing occlusion, shakiness, and tilt scores against empirically determined thresholds. The remaining videos are passed to the classification stage.

3. **Filtering:** In the filtering step, we determine videos that are unusable by comparing occlusion, shakiness,

[1] A third footnote, here. Let's make this a rather short one to see how it looks.

[2] A fourth, and last, footnote.

and tilt scores against empirically determined thresholds. The remaining videos are passed to the classification stage.

4. **Filtering:** In the filtering step, we determine videos that are unusable by comparing occlusion, shakiness, and tilt scores against empirically determined thresholds. The remaining videos are passed to the classification stage.

You can use whatever symbols, accented characters, or non-English characters you need anywhere in your document; you can find a complete list of what is available in the *LaTeX User's Guide*[?].

2.2 Math Equations

You may want to display math equations in three distinct styles: inline, numbered or non-numbered display. Each of the three are discussed in the next sections.

$$S^f = \alpha_1 S^q + \alpha_2 S^d$$

$$L^s = (1-\zeta)L^e + \zeta L^b \nu$$

$$S^d = \sum_{j=1}^{N^\nu} \nu_{ij} * \Delta_j$$

$$\nu_{ij} = Diff(I_i^c, I_j^h)$$

$$S^{im} = \frac{1}{255} \frac{1}{N^w * N^h} \sum_{x=1,}^{N_w} \sum_{y=1}^{N_h} I(x,y)$$

$$S^f = S^b \times S^d$$

$$S^{bp} = \begin{cases} 1 - N^b/(0.25 * N^i) & if N^b/(0.25 * N^i) < 1 \\ 0 & otherwise \end{cases}$$

$$S^{im} = \frac{1}{255} \sqrt{\frac{1}{N^w * N^h} \sum_{x=1,}^{N_w} \sum_{y=1}^{N_h} (I(x,y) - \bar{I})}$$

$$\mathcal{H} = \{S^b \times S^d | 1 \leq j \leq N^\nu\}$$

$$S^t = \frac{abs(\frac{1}{N^l}\sum_{i=1}^{N^l} o_i * l_i)}{\pi/4}$$

$$VS = \{\nu_{ij} | 1 \leq i \leq n; 1 \leq j \leq N^\nu; \forall i = j, \nu_{i,j} = 1\}$$

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

2.2.1 *Inline (In-text) Equations*

A formula that appears in the running text is called an inline or in-text formula. It is produced by the **math** environment, which can be invoked with the usual `\begin. . .\end` construction or with the short form `$. . .$`. You can use any of the symbols and structures, from α to ω, available in LaTeX[?]; this section will simply show a few examples of in-text equations in context. Notice how this equation: $\lim_{n\to\infty} x = 0$, set here in in-line math style, looks slightly different when set in display style. (See next section).

2.2.2 *Display Equations*

A numbered display equation – one set off by vertical space from the text and centered horizontally – is produced by the **equation** environment. An unnumbered display equation is produced by the **displaymath** environment.

Again, in either environment, you can use any of the symbols and structures available in LaTeX; this section will just give a couple of examples of display equations in context. First, consider the equation, shown as an inline equation above:

$$\lim_{n\to\infty} x = 0 \tag{1}$$

Notice how it is formatted somewhat differently in the **displaymath** environment. Now, we'll enter an unnumbered equation:

$$\sum_{i=0}^{\infty} x + 1$$

and follow it with another numbered equation:

$$\sum_{i=0}^{\infty} x_i = \int_0^{\pi+2} f \tag{2}$$

just to demonstrate LaTeX's able handling of numbering.

2.3 Citations

Citations to articles [?, ?, ?, ?], conference proceedings [?] or books [?, ?] listed in the Bibliography section of your article will occur throughout the text of your article. You should use BibTeX to automatically produce this bibliography; you simply need to insert one of several citation commands with a key of the item cited in the proper location in the `.tex` file [?]. The key is a short reference you invent to uniquely identify each work; in this sample document, the key is the first author's surname and a word from the title. This identifying key is included with each item in the `.bib` file for your article.

The details of the construction of the `.bib` file are beyond the scope of this sample document, but more information can be found in the *Author's Guide*, and exhaustive details in the *LaTeX User's Guide*[?].

This article shows only the plainest form of the citation command, using `\cite`. This is what is stipulated in the SIGS style specifications. No other citation format is endorsed or supported.

2.4 Tables

Because tables cannot be split across pages, the best placement for them is typically the top of the page nearest their initial cite. To ensure this proper "floating" placement of tables, use the environment **table** to enclose the table's contents and the table caption. The contents of the table itself

Table 2: Frequency of Special Characters

Non-English or Math	Frequency	Comments
Ø	1 in 1,000	For Swedish names
π	1 in 5	Common in math
$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage

Figure 1: A sample black and white graphic (.eps format).

must go in the **tabular** environment, to be aligned properly in rows and columns, with the desired horizontal and vertical rules. Again, detailed instructions on **tabular** material is found in the *LaTeX User's Guide*.

Immediately following this sentence is the point at which Table 1 is included in the input file; compare the placement of the table here with the table in the printed dvi output of this document.

To set a wider table, which takes up the whole width of the page's live area, use the environment **table*** to enclose the table's contents and the table caption. As with a single-column table, this wide table will "float" to a location deemed more desirable. Immediately following this sentence is the point at which Table 2 is included in the input file; again, it is instructive to compare the placement of the table here with the table in the printed dvi output of this document.

2.5 Figures

Like tables, figures cannot be split across pages; the best placement for them is typically the top or the bottom of the page nearest their initial cite. To ensure this proper "floating" placement of figures, use the environment **figure** to enclose the figure and its caption.

This sample document contains examples of **.eps** and **.ps** files to be displayable with LaTeX. More details on each of these is found in the *Author's Guide.*

As was the case with tables, you may want a figure that spans two columns. To do this, and still to ensure proper "floating" placement of tables, use the environment **figure*** to enclose the figure and its caption. and don't forget to end the environment with figure*, not figure!

Note that either **.ps** or **.eps** formats are used; use the `\epsfig` or `\psfig` commands as appropriate for the different file types.

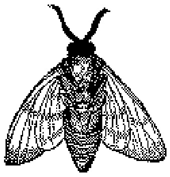


Figure 2: A sample black and white graphic (.eps format) that has been resized with the `epsfig` command.

2.6 Theorem-like Constructs

Other common constructs that may occur in your article are the forms for logical constructs like theorems, axioms, corollaries and proofs. There are two forms, one produced by the command `\newtheorem` and the other by the command `\newdef`; perhaps the clearest and easiest way to distinguish them is to compare the two in the output of this sample document:

This uses the **theorem** environment, created by the `\newtheorem` command:

THEOREM 1. *Let f be continuous on $[a,b]$. If G is an antiderivative for f on $[a,b]$, then*

$$\int_a^b f(t)dt = G(b) - G(a).$$

The other uses the **definition** environment, created by the `\newdef` command:

Definition 1. If z is irrational, then by e^z we mean the unique number which has logarithm z:

$$\log e^z = z$$

Two lists of constructs that use one of these forms is given in the *Author's Guidelines.*

There is one other similar construct environment, which is already set up for you; i.e. you must *not* use a `\newdef` command to create it: the **proof** environment. Here is a example of its use:

PROOF. Suppose on the contrary there exists a real number L such that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = L.$$

Then

$$l = \lim_{x\to c} f(x) = \lim_{x\to c} \left[gx \cdot \frac{f(x)}{g(x)} \right] = \lim_{x\to c} g(x) \cdot \lim_{x\to c} \frac{f(x)}{g(x)} = 0 \cdot L = 0,$$

which contradicts our assumption that $l \neq 0$. □

Complete rules about using these environments and using the two different creation commands are in the *Author's Guide*; please consult it for more detailed instructions. If you need to use another construct, not listed therein, which you want to have the same formatting as the Theorem or the Definition[**?**] shown above, use the `\newtheorem` or the `\newdef` command, respectively, to create it.

A *Caveat* for the TeX Expert

Because you have just been given permission to use the `\newdef` command to create a new form, you might think you can use TeX's `\def` to create a new command: *Please refrain from doing this!* Remember that your LaTeX source code is primarily intended to create camera-ready copy, but may be converted to other forms – e.g. HTML. If you inadvertently omit some or all of the `\def`s recompilation will be, to say the least, problematic.

3. CONCLUSIONS

This paragraph will end the body of this sample document. Remember that you might still have Acknowledgments or Appendices; brief samples of these follow. There is

Table 3: Some Typical Commands

Command	A Number	Comments
`\alignauthor`	100	Author alignment
`\numberofauthors`	200	Author enumeration
`\table`	300	For tables
`\table*`	400	For wider tables

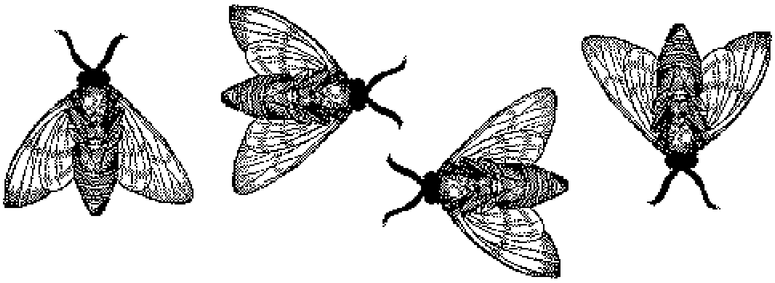


Figure 3: A sample black and white graphic (.eps format) that needs to span two columns of text.

still the Bibliography to deal with; and we will make a disclaimer about that here: with the exception of the reference to the LaTeX book, the citations in this paper are to articles which have nothing to do with the present subject and are used as examples only.

4. ACKNOWLEDGMENTS

This section is optional; it is a location for you to acknowledge grants, funding, editing assistance and what have you. In the present case, for example, the authors would like to thank Gerald Murray of ACM for his help in codifying this *Author's Guide* and the **.cls** and **.tex** files that it describes.

5. REFERENCES

APPENDIX

A. HEADINGS IN APPENDICES

The rules about hierarchical headings discussed above for the body of the article are different in the appendices. In the **appendix** environment, the command **section** is used to indicate the start of each Appendix, with alphabetic order designation (i.e. the first is A, the second B, etc.) and a title (if you include one). So, if you need hierarchical structure *within* an Appendix, start with **subsection** as the highest level. Here is an outline of the body of this document in Appendix-appropriate form:

A.1 Introduction

A.2 The Body of the Paper

A.2.1 Type Changes and Special Characters

A.2.2 Math Equations

Inline (In-text) Equations.

Display Equations.

A.2.3 Citations

A.2.4 Tables

A.2.5 Figures

A.2.6 Theorem-like Constructs

A Caveat for the TeX Expert

A.3 Conclusions

A.4 Acknowledgments

A.5 Additional Authors

This section is inserted by LaTeX; you do not insert it. You just add the names and information in the `\additionalauthors` command at the start of the document.

A.6 References

Generated by bibtex from your .bib file. Run latex, then bibtex, then latex twice (to resolve references) to create the .bbl file. Insert that .bbl file into the .tex source file and comment out the command `\thebibliography`.

B. MORE HELP FOR THE HARDY

The sig-alternate.cls file itself is chock-full of succinct and helpful comments. If you consider yourself a moderately experienced to expert user of LaTeX, you may find reading it useful but please remember not to change it.