Use of a 3D Graphic Environment in Superimposition Analysis

Tom Tolleson
Interaction Design laboratory
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
+1 919 9663611
ttolleso@email.unc.edu

Abstract

To support effective browsing and increased relevance of surrogates (smaller objects that describe a video), this project intends to interpret and represent video information and its relationship to superimpositions (subtitles or any other alpha-numeric information used to convey meaning) through a 3D graphic environment. This environment is intended to represent the overall structure of a video, or group of videos as well as to contrast individual videos or groups of videos.

Introduction

This project is an extension to a mark-up language that I created for the comic book Ranma 1/2. The markup language I created was a color-coding system to distinguish the various categories of alphanumeric information that existed outside of the visual channel in the comic book.
Upon later reviewing the color-coding system, I realized that a hierarchy could be discerned in the alphanumeric channel. The structure of this hierarchy was determined by the level/degree of sophistication of consciousness that informed each classification of alphanumeric information.

Interpreting Data for the Project

The classifications, along with their position in the hierarchy, as well as color code are as follows in ascending order:

1) Onomatopoeia sounds produced by movements of matter. Green.
2) Onomatopoeia sounds or other non-articulated sounds made by human voices. Blue.
3) Articulated verbal communication. Orange.
4) Written language. Brown.
5) Omniscient narration. Red.

All non-alphanumeric (visual information) is white. The alphanumeric information on each page was then filled with the color of its class in the hierarchy. Because comic books are a visual medium, it is assumed that the relative size of each incident of information is representative of its intended importance on the page. Subsequently, we can determine the overall intended importance of each classification of alphanumeric information in the entire book. Taking this one step further, if it were possible to index large numbers of volumes of comics, we could then begin to make comparisons between genres, series and cultures.

The Retrieval Process

The retrieval aspect of this project is not the main focus, but rather the ability of the user to browse retrieved documents and make comparisons. The project makes the following assumptions concerning retrieval:

1) All documents to be retrieved are indexed.
2) All queries result from a well-articulated need.
3) The smallest document size will be a chapter from a book. This smallest possible document is a story element that will be assumed to have a certain cohesive theme in information and be composed of enough contiguous pages to produce vectors that describe the content, by classifications, of the alphanumeric channel.

Thus, there is a fairly attenuated set of retrieval options and this tool is exclusive of users that are not attempting to retrieve multiple documents for the purposes of comparing the vectors that describe the document. The initial purpose of this is, in fact, to allow users to browse retrieved documents and make comparisons between retrieved documents based on content vectors.

These vectors themselves are a result of the indexing process and exist only because of the classifications assigned to them through the creation of a mark-up language for the documents. Thus the classifications, indexing and hierarchy are a result of a human filter interpreting the information and making judgments.

The Browsing Process

Browsing and comparison is the initial purpose of the tool. Once documents are retrieved, the vectors are created through a process of analyzing the overall space on each page dedicated to the incidents of alphanumeric channel information by their classification. This is done by assessing the total number of pixels on each scanned page of the document relative to the number of pixels on each page that are occupied by alphanumeric information. Pages 2, 3 and 4 of Ranma 1/2 (Appendix A) are assessed in table 1

	Sounds caused by matter	Human noise	Articulated human speech	Written language	Omniscient narrator	
Page 2	110994	0	46642	33582	0	
Page 3	109716	0	99178	6973	50975	
Page 4	83941	38915	84575	0	0	
						Total possible
						676,564

Table 1

The total number of pixels per page is 676,564. The number of pixels occupied by each classification on each page appears underneath its classification. The hierarchy begins on the left with its "lowest" level being "Sounds caused by matter" and spanning to the "Omniscient narrator" on the far right. This information is then used to create a two-dimensional image representing the page, such as page 2 (figure 1).
The indexing hierarchy informs the organization of the elements of the two-dimensional image. Creation of the two-dimensional image begins with a white background proportional to the total number of pixels on each page. Then blocks of color are added, one on top of the other in the center of the image. The blocks are the same color and size of each classification in the system based on the hierarchy moving from left to right in table 1 . We see in table 1 that we have 110994 pixels on page 2 occupied by alphanumeric information categorized as "sounds caused by matter" and the color for that classification is green.
We must then determine the amount of space that this element of the image should occupy by dividing 676,564 by 110,994.

676,564/110,994 = 6.095

So we can see that roughly 1/6 of the white area should be occupied by a green block (sounds caused by matter); 1/14 should contain an orange block (articulated human speech); finally, approximately 1/20 should contain a brown block (written language).

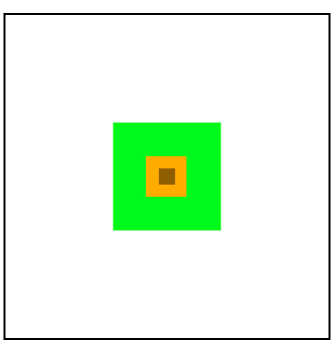


Figure 1

The Third Dimension

This image is then applied to a two-dimensional Cartesian plane (figure 2) comprised of axes x and z. The computer then interprets the third dimension that occurs along axis y based on the color information in the image. Movement along the y axis is caused by the presence of information in the alphanumeric channel at different levels of the hierarchy in table 1.

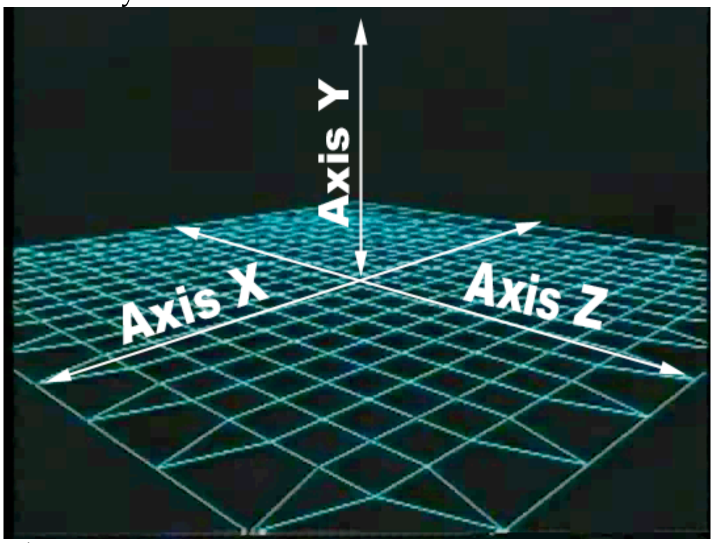


Figure 2

Thus green information causes a deformation in the grid rising to position 1 on the y axis due to it being the first position in the hierarchy. Orange will cause a deformation rising to position 3 on the y axis as it is the third level in the hierarchy. The deformations occur only on the portion of the grid defined by each color in the original two-dimensional image (figure 1).

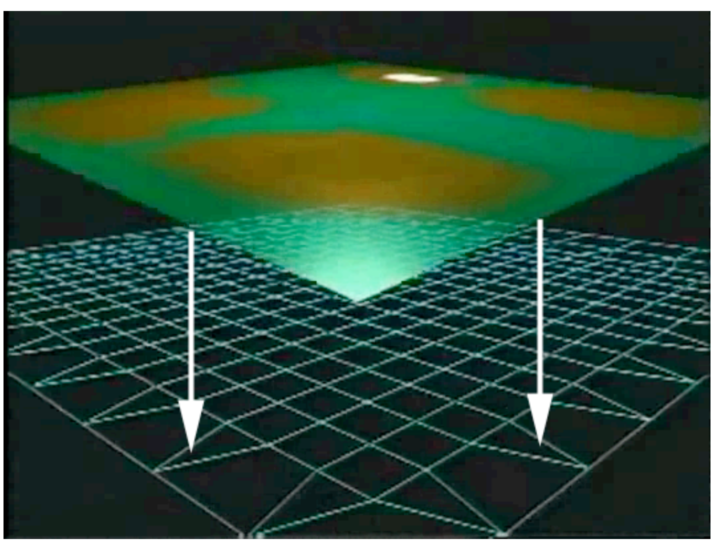


Figure 2.1

Deformations are then created that resemble topographical gradients (figure 3).

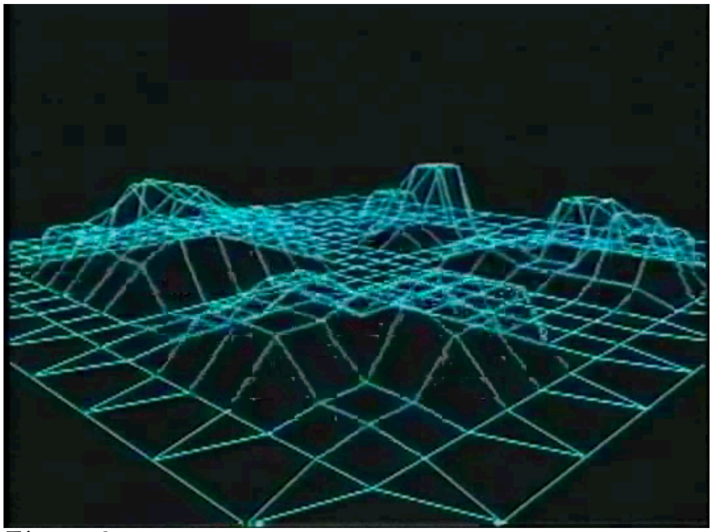
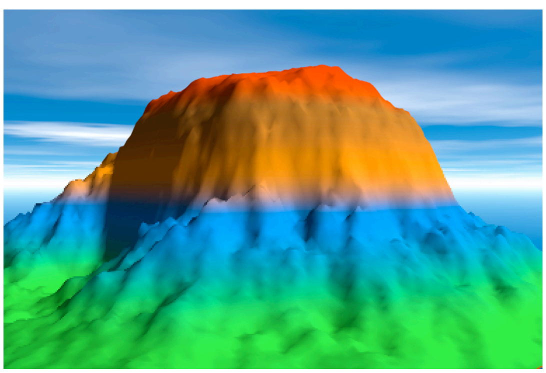
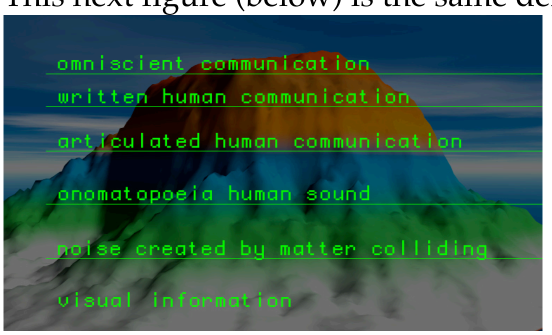


Figure 3

With the color-coding system of the indexing hierarchy applied to a sample deformation we see the below results.

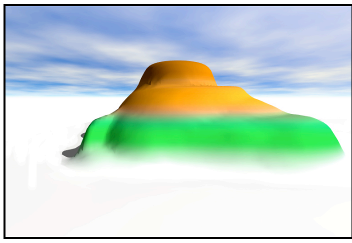


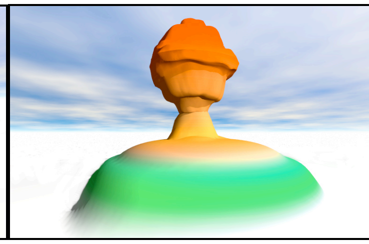
This next figure (below) is the same deformation with a key overlay.

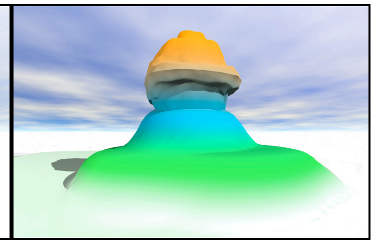


Application to the comic book pages

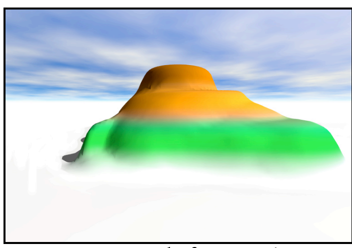
The three pages of Ranma 1/2 (Appendix A) were then converted into three-dimensional deformations through this process. We see the results below.

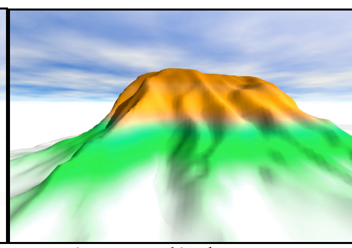

Page 2

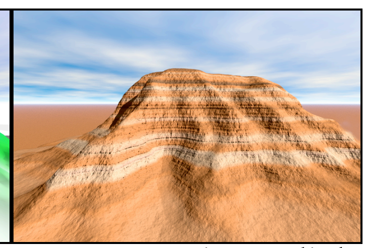

Page 3


Page 4

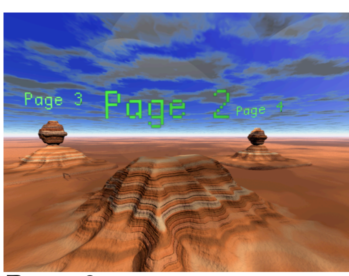
An erosive modeling algorithm is then applied to the deformations. After the application of the algorithm, a natural texture-mapping algorithm is added. Below we see these two final transitions as applied to page 2.


Page 2 as a deformation

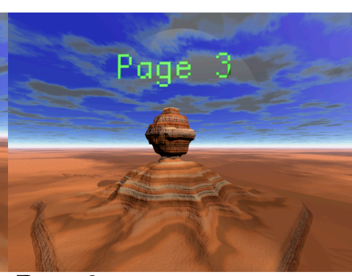

Erosion applied


Texture-mapping applied

So now we have a very natural looking topographical deformation that is created. When these algorithms are applied to all three pages we see the following results:



Page 2



Page 3

Page 4

Use of the tool

This tool is intended to be an enhancement for browsing retrieved documents. The final result of the project will be a completely immersive three-dimensional environment where the user can identify and understand shapes with the assistance of a heads-up-display similar to those found in the cockpits of military fighter aircraft (figure4) or in geographic DEM data (figure 5).

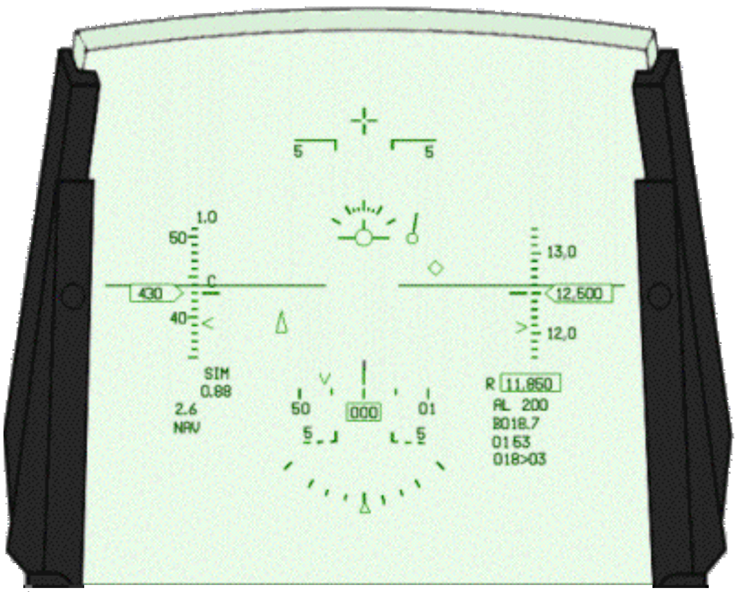


Figure 4

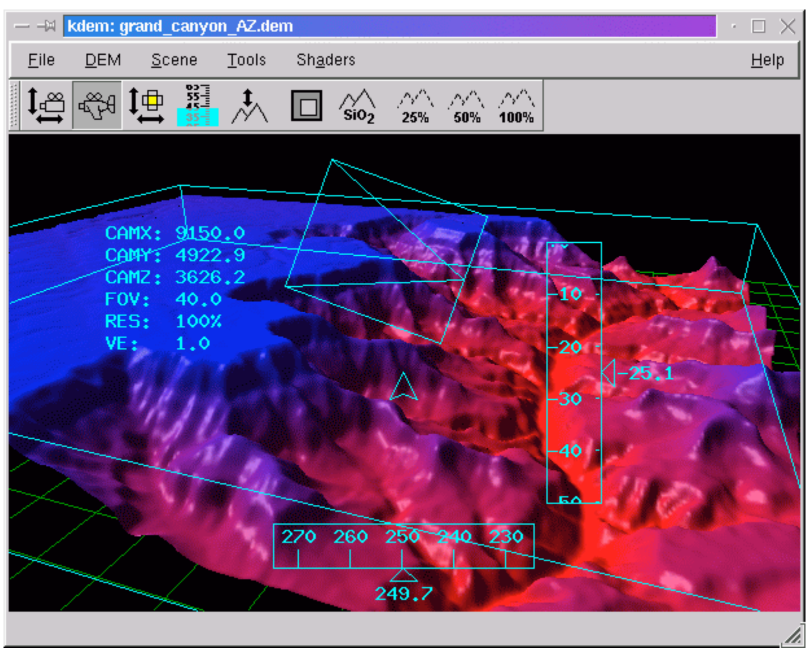


Figure 5

The purpose of the creation of topographical structures is based on some biometric findings and anthropological theories. Companies such as Real User Corporation have been utilizing biometrics as a replacement for alphanumeric situations such as passwords. Instead of an alphanumeric password, a user's password can be a series of human faces. The user doesn't need to remember anything about the faces, they just have to be familiar with them. Users seem to be able to retain the information (or familiarity) with the faces more easily than remembering a password. Randomly generated passwords often demand some sort of mnemonic trick by the user to aid in remembering his/her password.

Why can people more easily gain familiarity with faces as opposed to passwords? One possibility lies in the innate ability we have to perceive faces "Nine minutes after birth, newborns already prefer faces to other visual patterns and prefer to look at a face-like pattern rather than at a scrambled face." (Schwarzer & Leder55).

People use completely different regions of their brain to retain facial information than those that are used to remember alphanumeric information. "A number of separate studies have shown that there are substantial numbers of cells in the superior temporal sulcus of the macaque temporal cortex that respond preferentially to faces and show little if any response to other simple objects (such as clocks) or arousing objects (such as snakes." (Bruce 8).

This modular view of the brain includes other regions that seem to have evolved to respond to certain types of environmental information. Many anthropologists believe that the parietal lobe may be key to our understanding of landmarks and location in the world around us. For many years, forensic anthropologists used a simple brain-to-body-mass threshold that determined if a fossil specimen was to be classified as ape or human. Any specimen determined to have a brain size above 700 cc was determined to be human. However an anthropologist named Ralph Holloway began making molds of brains from the skull-cases of fossil specimens and was able to determine the rough organization of the specimens' brains. This discovery had a significant impact on the field because he was able to demonstrate that not only did the brains of primates grow as they evolved into modern humans, but their organization altered as well (Leakey 197).

The occipital lobe, which primarily processes information from the eyes, became less prominent and the parietal lobe, which part of the super-association region, began to grow. "Forgetting for a minute their relative size, one can distinguish between a chimpanzee's brain and a human's by the shape of the cerebral jigsaw: basically, a brain with small temporal and parietal lobes and relatively large occipital lobes is apelike; by contrast, a brain with a human organization has large temporal and parietal lobes, and relatively smaller occipital lobes" (Leakey, 197).

This change is estimated to have occurred around 3 million years ago, when the last of the Australopithecus boisei were seen and their contemporary, Homo Habilis was first appearing. Homo Habilis is one of our earliest ancestors that can be unmistakably linked to the path of becoming human. "What shaped the human brain from its apelike origins would seem to be a matter of the social responses to the new ecological niche that was being carved out by our ancestors: a tighter, more complex social grouping, bringing with it a greater need to manipulate the ramifying relationships, to behave cooperatively, and to be more aware of the surrounding terrain would all have exerted keen selective pressures" (Leakey 198).

So if one of our brain-defining moments was our climb down from the trees and resulting journey out into the savannahs of Africa; if the occipital lobe began to shrink with our need for grasping and climbing, and our parietal lobe began to grow with our nomadic need to understand terrain and remember landmarks, then our ability to quickly and innately understand terrains and form cognitive maps is as innate as our ability to recognize faces. And if this is true, then an immersive terrain could more effectively speak to a user than volumes of alphanumeric data in spreadsheets.

This tool is in no way intended to replace any existing browse features or data structures that are available in information retrieval. However, it may create an alternate set of visual stimulation able to assist users in quickly comprehending the information they have retrieved. The examination of results in a search is an important step in the retrieval process. This tool might help a user quickly assess large amounts of retrieved information by using his/her innate skills in determining a terrain that is an abstraction of indexed vectors.

Works Cited

Vicki Bruce, Face Recognition (Lawrence Erlbaum Associates Ltd., 1991)

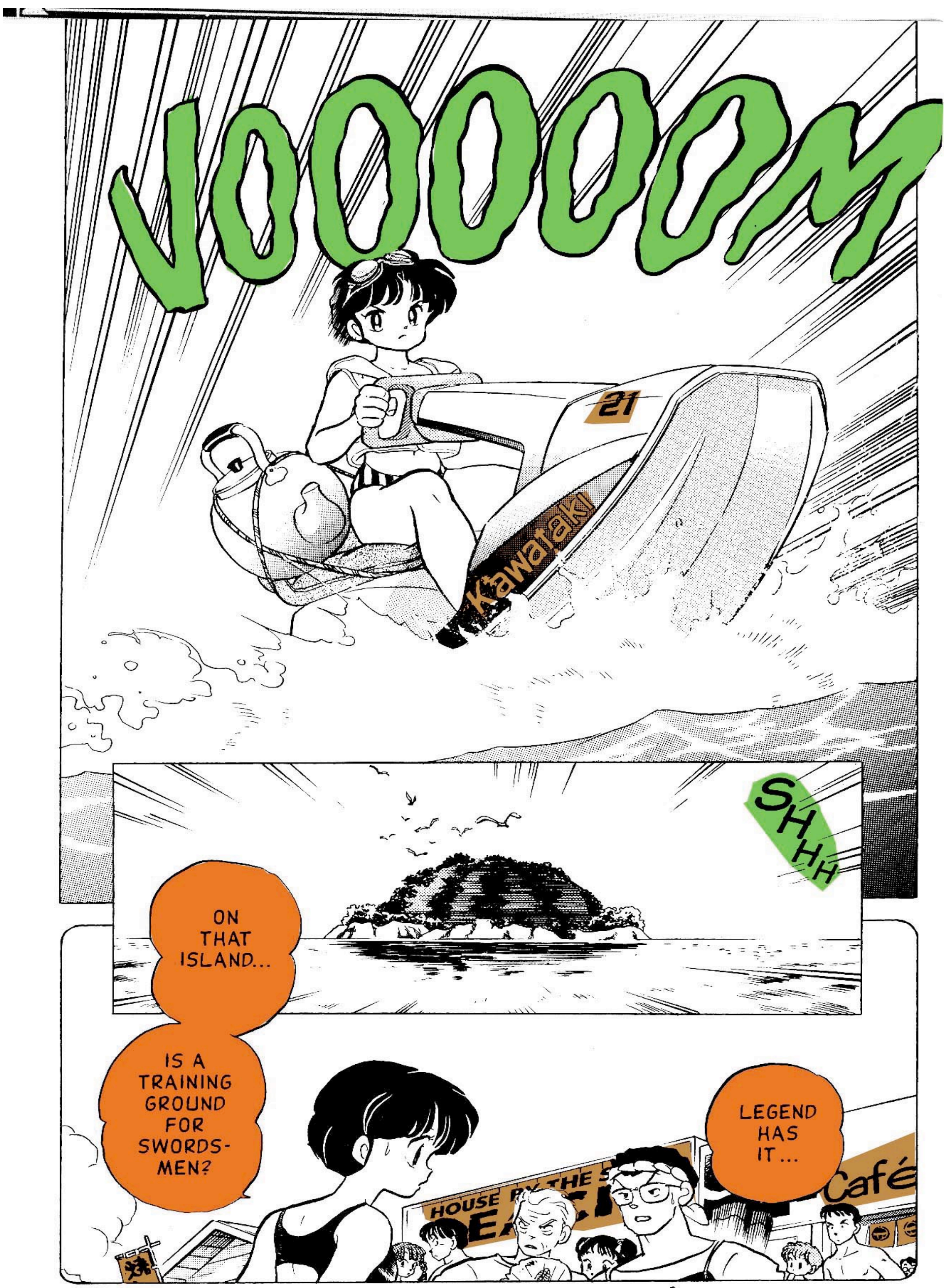
Richard E. Leakey, Origins (E.P. Dutton, New York, June 1978)

Gudrun Schwarzer & Helmut Leder, The Development of Face Processing (Hogrefe & Huber Publishers, 2003)

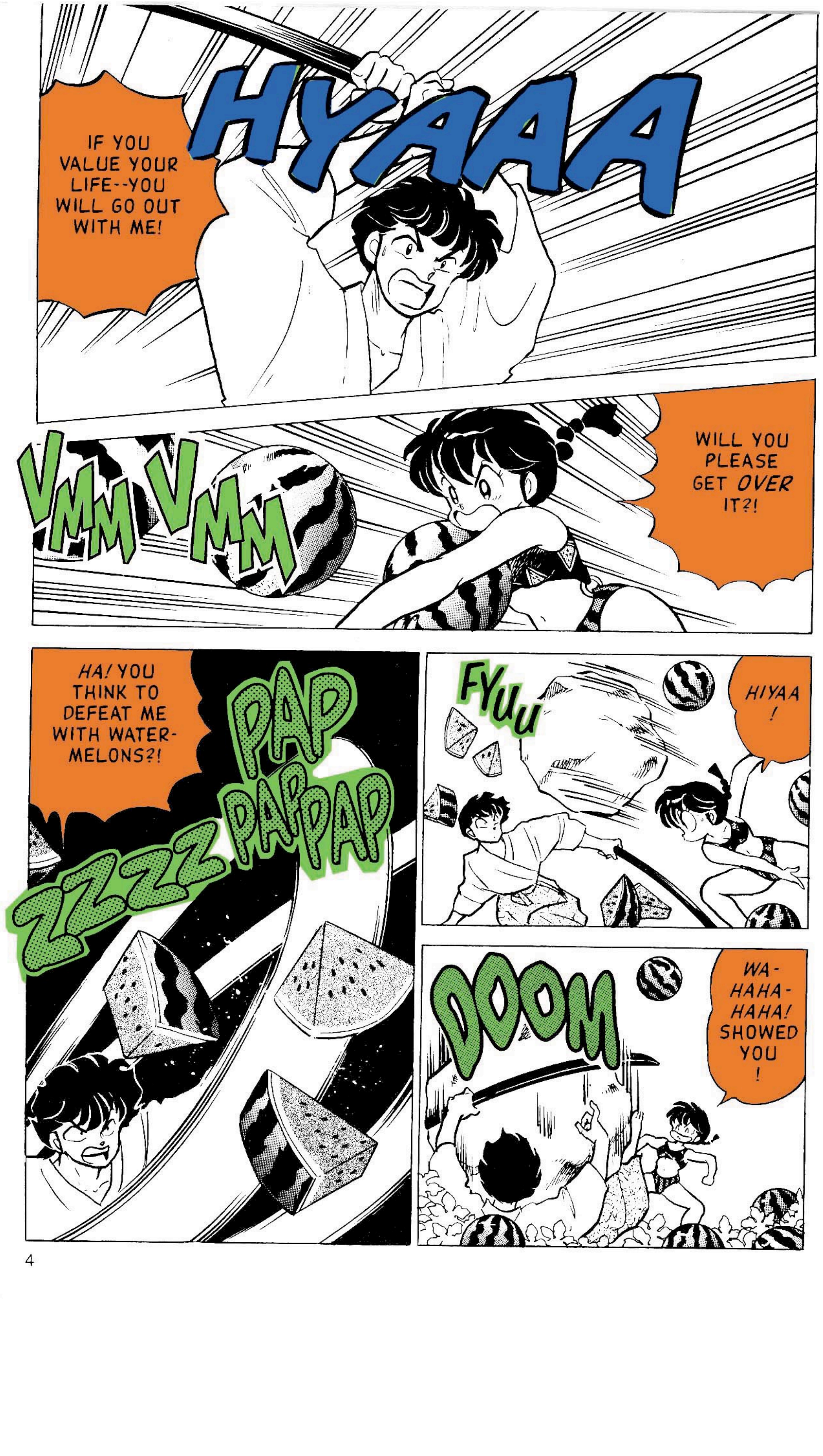
Rumiko Takahashi, Ranma 1/2 (Viz Communications, 2001)

Appendix A

Ranma 1/2


VOOOOOM
21
Kawataki
SHHH
ON THAT ISLAND...
IS A TRAINING GROUND FOR SWORDSMEN?
LEGEND HAS IT...
HOUSE BY THE S
Café


...THAT THE TRAINING IS SO RIGOROUS, SO TERRIFYING, THAT MANY OF THOSE WHO ATTEMPT IT COME BACK WITH AMNESIA.
KUNO WENT THERE TO TRAIN FOR HIS FIGHT AGAINST RANMA...
THERE'S NO DOUBT ABOUT IT...
VOOOOOOM
IN EXCHANGE FOR HIS MEMORIES...
AND ON WATERMELON ISLAND...
RANMA IS IN DANGER !
ARRGH!!
HEH HEH HEH! I'VE FOUND YOU AT LAST....
HE GAINED MASTERY OF A DEADLY TECHNIQUE.


IF YOU VALUE YOUR LIFE--YOU WILL GO OUT WITH ME!
HYAAA
VMM VMM
WILL YOU PLEASE GET *OVER* IT?!
HA! YOU THINK TO DEFEAT ME WITH WATER-MELONS?!
PAP PAP PAP PAP
ZZZZ
FYUU
HIYAA !
DOOM
WA-HAHA-HAHA! SHOWED YOU !