Content-Based Information Retrieval and Digital Libraries

Gary (Gang) Wan and Zao Liu

This paper discusses the applications and importance of content-based information retrieval technology in digital libraries. It generalizes the process and analyzes current examples in four areas of the technology. Content-based information retrieval has been shown to be an effective way to search for the type of multimedia documents that are increasingly stored in digital libraries. As a good complement to traditional text-based information retrieval technology, content-based information retrieval will be a significant trend for the development of digital libraries.

With several decades of their development, digital libraries are no longer a myth. In fact, some general digital libraries such as the National Science Digital Library (NSDL) and the Internet Public Library are widely known and used. The advance of computer technology makes it possible to include a colossal amount of information in various formats in a digital library. In addition to traditional text-based documents such as books and articles, other types of materials—including images, audio, and video—can also be easily digitized and stored. Therefore, how to retrieve and present this multimedia information effectively through the interface of a digital library becomes a significant research topic.

Currently, there are three methods of retrieving information in a digital library. The first and the easiest way is free browsing. By this means, a user browses through a collection and looks for desired information. The second method—the most popular technique used today—is text-based retrieval. Through this method, textual information (full text of text-based documents and/or metadata of multimedia documents) is indexed so that a user can search the digital library by using keywords or controlled terms. The third method is content-based retrieval, which enables a user to search multimedia information in terms of the actual content of image, audio, or video (Marques and Furht 2002). Some content features that have been studied so far include color, texture, size, shape, motion, and pitch.

While some may argue that text-based retrieval techniques are good enough to locate desired multimedia information, as long as it is assigned proper metadata or tags, words are not sufficient to describe what is sometimes in a human's mind. Imagine a few examples: A patron comes to a public library with a picture of a rare insect. Without expertise in entomology, the librarian won't know where to start if only a text-based information retrieval system is available. However, with the help of content-based image retrieval, the librarian can upload the digitized image of the insect to an online digital image library of insects, and the system will retrieve similar images with detailed description of this insect. Similarly, a patron has a segment of music audio, about which he or she knows nothing but wants to find out more. By using the content-based audio retrieval system, the patron can get similar audio clips with detailed information from a digital music library, and then listen to them to find an exact match. This procedure will be much easier than doing a search on a text-based music search system. It is definitely helpful if a user can search this non-textual information by styles and features.

In addition, the advance of the World Wide Web brings some new challenges to traditional text-based information retrieval. While today's Web-based digital libraries can be accessed around the world, users with different language and cultural backgrounds may not be able to do effective keyword searches of these libraries. Content-based information retrieval techniques will increase the accessibility of these digital libraries greatly, and this is probably a major reason it has become a hot research area in the past decade. Ideally, a content-based information retrieval system can understand the multimedia data semantically, such as its objects and categories to which it belongs. Therefore, a user is able to submit semantic queries and retrieve matched results. However, a great difficulty in the current computer technology is to extract high-level or semantic features of multimedia information. Most projects still focus on lower-level features, such as color, texture, and shape.

Simply put, a typical content-based information retrieval system works in this way: First, for each multimedia file in the database, certain feature information (e.g., color, motion, or pitch) is extracted, indexed, and stored. Second, when a user composes a query, the feature information of the query is calculated as vectors. Finally, the system compares the similarity between the feature vectors of the query and multimedia data, and retrieves the best matching records. If the user is not satisfied with the retrieved records, he or she can refine the search results by selecting the most relevant ones to the search query, and repeat the search with the new information. This process is illustrated in figure 1.

The following sections will examine some existing content-based information retrieval techniques for most common information formats (image, audio, and video) in digital libraries, as well as their limitations and trends.

Gary (Gang) Wan (gwan@tamu.edu) is a Science Librarian and Assistant Professor, and **Zao Liu** (zliu@tamu.edu) is a Distance Learning Librarian and Assistant Professor at Sterling C. Evans Library, Texas A&M University, College Station, Texas.

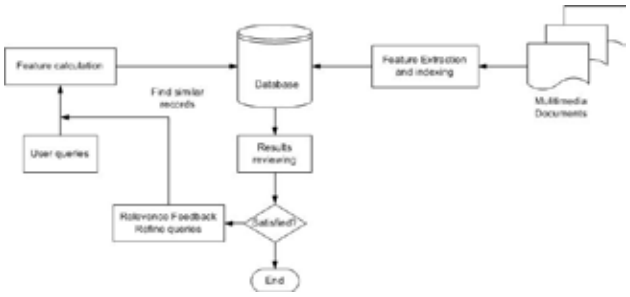

Figure 1. The general process of content-based information retrieval

Content-based image retrieval

There have been a large number of different content-based image retrieval (CBIR) systems proposed in the last few years, either building on prior work or exploring novel directions. One similarity among these systems is that most perform feature extraction as the first step in the process, obtaining global image features such as color, shape, and texture (Datta et al., 2005).

One of the most well-known CBIR systems is query by image content (QBIC), which was developed by IBM. It uses several different features, including color, sketches, texture, shape, and example images to retrieve images from image and video databases. Since its launch in 1995, the QBIC model has been employed for quite a few digital libraries or collections. One recent adopter is the State Hermitage Museum in Russia (www.hermitage.ru), which uses QBIC for its Web-based digital image collection. Users can find artwork images by selecting colors from a palette or by sketching shapes on a canvas. The user can also refine existing search results by requesting all artwork images with similar visual attributes. The following screenshots demonstrate how a user can do a content-based image search with QBIC technology.

In figure 2.1, the user chooses a color from the palette and composes the color schema of artwork he or she is looking for. Figure 2.2 shows the artwork images that match the query schema.

Another example of digital libraries or collections that have incorporated CBIR technology is the National Science Foundation's International Digital Library Project (www.memorynet.org), a project that is composed of several image collections. The information retrieval system for these collections includes both a traditional text-based search engine and a CBIR system called SIMPLIcity (Semantics-sensitive Integrated Matching for Picture Libraries) developed by Wang et al. (2001) of Pennsylvania State University.

From the front page of these image collections, a user can choose to display a random group of images (figure

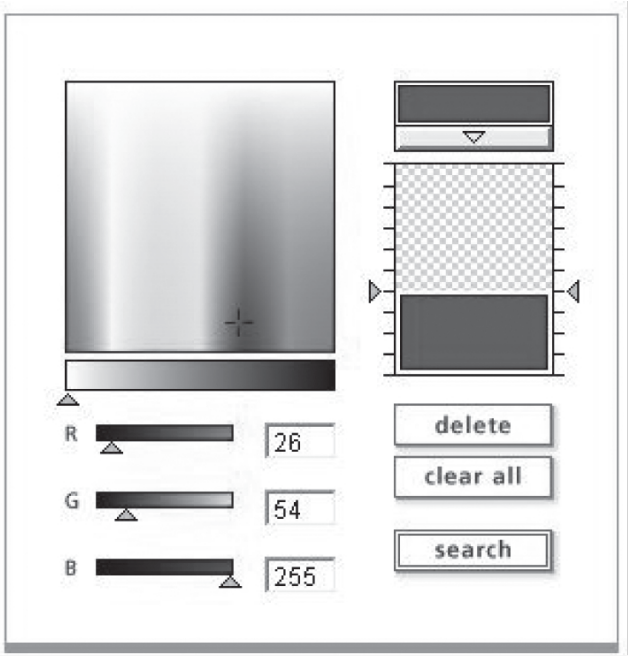

Figure 2.1. A user query

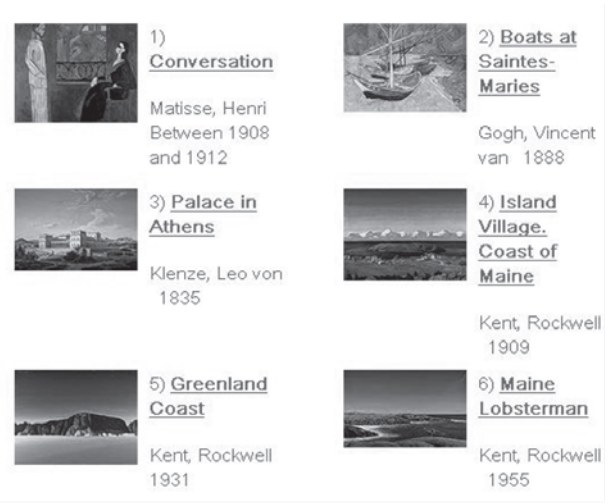

Figure 2.2. The search results for this query

3.1). Below each image is a "similar" button; clicking this allows the user to view a group of images that contain similar objects to the previously selected one (figure 3.2). By providing feedback to the search engine this way, the user can find images of desired objects without knowing their names or descriptions.

Simply put, SIMPLIcity segments each image into small regions, extracts several features (such as color,

location, and shape) from these small regions, and classifies these regions into some semantic categories (such as textured/nontextured and graph/photograph). When computing the similarity between the query image and images in the database, all these features will be considered and integrated, and best matching results will be retrieved (Wang et al., 2001).

Similar applications of CBIR technology in digital libraries include the University of California–Berkeley's Digital Library Project (http://bnhm.berkeley.edu), the National STEM Digital Library (ongoing), and Virginia Tech's anthropology digital library, ETANA (ongoing).

While these feature-based approaches have been explored over the years, an emerging new research direction in CBIR is automatic concept recognition and annotation. Ideally, automatic concept recognition and annotation can discover the concepts that an image conveys and assign a set of metadata to it, thus allowing image search through the use of text. A trusted automatic concept recognition and annotation system can be a good solution for large data sets. However, the semantic gap between computer processors and human brains remains the major challenge in the development of a robust automatic concept recognition and annotation system (Datta et al., 2005). A recent example of efforts in this field is Li and Wang's ALIPR (Automatic Linguistic Indexing of Pictures—Real Time, http://alipr.com) project (2006). Through a Web interface, users are able to search images in several different ways: They may do text searches and provide feedback to the system to find similar images. Users may also upload an image, and the system will perform concept analysis and generate a set of annotations or tags automatically, as shown in figure 4. The system then retrieves images from the database that are visually similar to the uploaded image. In the process of automatic annotation, if the user doesn't think the tags given by the system are suitable, he or she can input other tags to describe the image. This is also the "training" process for the ALIPR system.

Since CBIR is the major research area and has the longest history in content-based information retrieval, there are many models, products, and ongoing projects in addition to the above examples. As image collections become a significant part of digital libraries, more attention has been paid to possibilities of providing content-based image search as a complement to existing metadata search.

Figure 3.1. A group of random images in the collection

Figure 3.2. CBIR results

Content-based audio retrieval

Compared with CBIR, content-based audio retrieval (CBAR) is relatively new, and fewer research projects on it can be found. In general, existing CBAR approaches start from the content analysis of audio clips. An example of this content analysis is extracting basic audio elements, such as duration, pitch, amplitude, brightness, and band-

Figure 4. ALIPR's automatic annotation feature

width (Wold et al., 1996).

Because of the great difficulties in recognizing audio content, research in this area is less mature than that in content-based image and video retrieval. Although no CBAR system has been found to be implemented by any digital library so far, quite a few projects provide good prototypes or directions.

One good example is Zhang and Kuo's (2001) research project on audio classification and retrieval. The prototype system is composed of three stages: coarse-level audio segmentation, fine-level classification, and audio retrieval. In the first stage, audio signals are semantically segmented and classified into several basic types including speech, music, song, speech with music background, environment sounds, and silence. Some physical audio features—such as the energy function, the fundamental frequency, and the spectral peak tracks—are examined in this stage. In the second stage, further classification is conducted for every basic type. Features are extracted from the time-frequency representation of audio signals to reveal subtle differences of timbre and pattern among different classes of sounds. Based on these differences, the coarse-level segmentation obtained in stage one can be classified to narrower categories. For example, speech can be differentiated into the voices of men, women, and children. Finally, in the information retrieval stage, two approaches—query-by-keyword and query-by-example—are employed. The query-by-keyword approach is more like the traditional text-based search system. The query-by-example approach is similar to content-based image retrieval systems where an image can be searched by color, texture, and histogram, and audio clips can be retrieved with distinct features, such as timbre, pitch, and rhythm. This way, a user may choose from a given list of features, listen to the retrieved samples, and modify the input feature set to get more desired results. Zhang and Kuo's prototype is a very typical and classic CBAR system. It is relatively mature and can be used by large digital audio libraries.

More recently, Li et al. (2003) proposed a new feature extraction method particularly for music genre classification named Daubechies Wavelet Coefficient Histograms (DWCHs). DWCHs capture the local and global information of music signals simultaneously by computing their histograms. Similar to other CBAR strategies, this method divides the process of music genre classification into two steps: feature extraction and multi-class classification. The music signal information representing the music is extracted first, and then an algorithm is used to identify the labels from the representation of the music sounds with respect to their features.

Since the decomposition of audio signal can produce a set of subband signals at different frequencies corresponding to different characteristics, Li et al. (2003) proposed a new methodology, the DWCHs algorithm, for feature extraction. With this algorithm, the decomposition of the music signals is obtained at the beginning, and then a histogram of each subband is constructed. Hence, the energy for each subband is computed, and the characteristics of the music are represented by these subbands. One finding from this research reveals that this methodology, along with advanced machine learning techniques, has significantly improved accuracy of music genre classification (Li et al. 2003). Therefore, this methodology potentially can be used by those digital music libraries widely developed in past several years.

Content-based video retrieval

Content-based video retrieval (CBVR) is a more recent research topic than CBIR and CBAR, partly because the digitization technology for video appeared later than those for image and audio. As digital video Websites such as YouTube and Google Video become more popular, how to retrieve desired video clips effectively is a great concern. Searching by some features of video, such as motion and texture, can be a good complement to the traditional text-based search method.

One of the earliest examples is the VideoQ system developed by Chang et al. (1997) of Columbia University (www.ctr.columbia.edu/VideoQ), which allows a user to search video based on a rich set of visual features and spatio-temporal relationships. The video clips in the database are stored as MPEG files. Through a Web interface, the user can formulate a query scene as a collection of objects with different attributes, including motion, shape, color, and texture. Once the user has formulated the query, it is sent to a query server, which contains several databases for different content features. On the query server, the similarities between the features of each object specified in the query and those of the objects in the database are computed; a list of video clips is then retrieved based on their similarity values. For each of these video clips, key-frames are dynamically extracted from the video database and returned to browser. The matched objects are highlighted in the returned key-frame. The user can interactively view these matched video clips by simply clicking on the key-frame. Meanwhile, the video clip corresponding to that key-frame is extracted from the video database (Chang et al. 1997). Figures 5.1–5.2 show an example of a visual search through the VideoQ system.

Many other CBVR projects also examine these content features and try to find more efficient ways to retrieve data. A recent example is Wang et al.'s (2006) project, Vferret, a content-based similarity search tool for continuous archived video. The Vferret system segments video data into clips and extracts both visual and audio features as metadata. Then a user can do a metadata search or

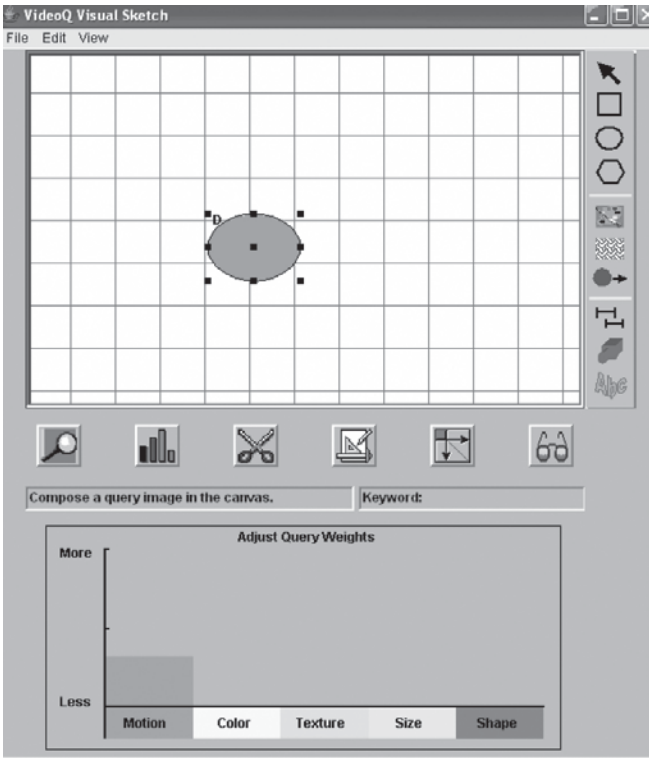

Figure 5.1. The user composes a query

Figure 5.2. Search results for the sample query

content-based search to retrieve desired video clips. In the first stage, a simple segmentation method is used to split the archived digital video into five-minute video clips. The system then extracts twenty image frames evenly from each of these five-minute video clips for visual feature extraction. Additionally, the system splits the audio channel of each clip into twenty individual fifteen-second segments for further audio feature extraction. In the second stage, both audio and visual features are extracted. For visual features, the color element is used as the content feature. For audio features, 154 audio features originally used by Ellis and Lee (2004) to describe audio segments are computed. For each fifteen-second video segment, the visual feature vector extracted from the sample image and the audio feature vector extracted from the corresponding audio segment are combined into a single feature vector. In the information retrieval stage, the user submits a video clip query at first, then its feature vector is computed and compared with that of video clips in the database, and the most similar clips are retrieved (Wang et al. 2006).

Similar projects in this area include Carnegie Mellon University's Informedia Digital Video Library (www.informedia.cs.cmu.edu) and MUVIS of Finland's Tampere University of Technology (http://muvis.cs.tut.fi/index.html).

Content-based information retrieval for other digital formats

With the advance of digitization technology, the content and formats of digital libraries are much richer than before. They are not limited to text, image, audio, and video. Some new formats of digital content are emerging. Digital libraries of 3-D objects are good examples.

Since 3-D models have arbitrary topologies and cannot be easily "parameterized" using a standard template as in the case for 2-D forms (Bustos et al. 2005), content-based 3-D model retrieval is a more challenging research topic than other multimedia formats discussed earlier. So far, four types of solutions—primitive-based, statistics-based, geometry-based, and view-based—have been found (Bimbo and Pala 2006). Primitive-based solutions represent 3-D objects with a basic set of parameterized primitive elements. Parameters are used to control the shape of each primitive element and to fit each primitive element with a part of the model. With statistics-based approaches, shape descriptions based on statistical mod-

els are created and measured. Geometry-based methods, however, use geometric properties of the 3-D object and their measures as global shape descriptors. For view-based solutions, a set of 2-D views of the model and descriptors of their content are used to represent the 3-D object shape (Bimbo and Pala 2006).

Another novel example is Moustakas et al.'s (2005) project on 3-D model search using sketches. In the experimental system, the vector of geometrical descriptors for each 3-D model is calculated during the feature extraction stage. In the retrieval stage, a user can initially use one of the sketching interfaces (such as the virtual reality interface or by using an air mouse) to sketch a 2-D contour of the desired 3-D object. The 2-D shape is recognized by the system, and a sample primitive is automatically inserted in the scene. Next, the user defines other elements that cannot be described by the 2-D contour, such as the height of the object, and manipulates the 2-D contour until it reaches its target position. The final query is formed after all the primitives are inserted. Finally, the system computes the similarities between the query model and each 3-D model in the database, and renders the best matching records.

An online demonstration can be found for a European project specifically designed for a 3-D digital museum collection, SCULPTEUR (www.sculpteurweb.org). From its Web-based search interface, a user can choose to do a metadata search or content-based search for a 3-D object. The search strategy here is somewhat similar to that in some CBIR systems: the user can upload a 3-D model in VRML formats, then select a search algorithm (such as similar color, texture, etc.) to perform a search within a digital collection of 3-D models. As 3-D computer visualization has been widely used in a variety of areas, there are more research projects focusing on the content-based information retrieval techniques for this new multimedia format.

Conclusion

There is no doubt that content-based information retrieval technology is an emerging trend for digital library development and will be an important complement to the traditional text-based retrieval technology. The ideal CBIR system can semantically understand the information in a digital library, and render users the most desirable data. However, the machine understanding of semantic information still remains to be a great difficulty. Therefore, most current research projects, including those discussed in this paper, deal with the understanding and retrieval of lower-level features or physical features of multimedia content. Certainly, as related disciplines such as computer vision and artificial intelligence keep developing, more researches will be done on higher-level feature-based retrieval.

In addition, the growing varieties of multimedia content in digital libraries have also brought many new challenges. For instance, 3-D models now become important components of many digital libraries and museums. Content-based retrieval technology can be a good direction for this type of content, since the shapes of these 3-D objects are often found more effectively if the user can compose the query visually. New CBIR approaches need to be developed for these novel formats.

Furthermore, most CBIR projects today tend to be Web-based. By contrast, many project were based on client applications in the 1990s. These Web-based CBIR tools will have significant influence on digital libraries or repositories, as most of them are also Web-based. Particularly in the age of Web 2.0, some large digital repositories—such as Flickr for images and YouTube and Google Video for video—are changing people's daily lives. The implementation of CBIR will be a great benefit to millions of users.

Since the nature of CBIR is to provide better search aids to end users, it is extremely important to focus on the actual user's needs and how well the user can use these new search tools. It is surprising to find that little usability testing has been done for most CBIR projects. Such testing should be incorporated into future CBIR research before it is widely adopted.

Bibliography

Bimbo, A. and P. Pala. 2006. Content-based retrieval of 3-D models. *ACM Transactions on Multimedia Computing, Communications, and Applications* 2, no. 1: 20–43.

Bustos, B., et al. 2005. Feature-based similarity search in 3-D object databases. *ACM Computing Surveys* 37, no. 4: 345–387.

Chang, S., et al. 1997). VideoQ: an automated content based video search system using visual cues. In *Proceedings of the 5th ACM International Conference on Multimedia,* E. P. Glinert, et al., eds. New York: ACM.

Datta R., et al. 2005. Content-based image retrieval: approaches and trends of the new age. In *Proceedings of the 7th International Workshop on Multimedia Information Retrieval, in Conjunction with ACM International Conference on Multimedia*, H. Zhang, , J. Smith, and Q. Tian, eds. New York: ACM.

Ellis, D. and K. Lee. Minimal-impact audio-based personal archives. In *Proceedings of the 1st ACM workshop on Continuous Archival and Retrieval of Personal Experiences CARPE*, J. Gemmell, et al., eds. New York: ACM.

Li, T., et al. 2003. A comparative study on content-based music genre classification. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, C. Clarke, et al., eds. New York: ACM.

Li, J. and J. Wang, J. 2006. Real-time computerized annotation of pictures. In *Proceedings of the 14th Annual ACM International*

Conference on Multimedia, K. Nahrstedt, et al., eds. New York: ACM.
Marques, O. and B. Furht. 2002. Content-based Image and Video Retrieval. Norwell, Mass: Kluwer.
Moustakas, K., et al. 2005. MASTER-PIECE: A multimodal (gesture+speech) interface for 3D model search and retrieval integrated in a virtual assembly application. *Proceedings of the eNTERFACE:* 62–75.
Wang, J., et al. 2001. SIMPLIcity: semantics-sensitive integrated matching for picture libraries. *IEEE Trans. Pattern Analysis and Machine Intelligence* 23, no. 9: 947–963.
Wang, Z., et al. 2006. VFerret: content-based similarity search tool for continuous archived video. In *Proceedings of the 3rd ACM Workshop on Continuous Archival and Retrival of Personal Experiences,* K. Maze et al., eds. New York: ACM.
Wold, E., et al. 1996. Content-based classification, search, and retrieval of audio. *IEEE MultiMedia* 3, no. 3: 27–36.
Zhang, T. and C. Kuo. 2001. *Content-based Audio Classification and Retrieval for Audiovisual Data Parsing*. Norwell, Mass.: Kluwer.

Index to Advertisers

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION

Information Technology and Libraries, Publication No. 280-800, is published quarterly in March, June, September, and December by the Library Information and Technology Association, American Library Association, 50 E. Huron St., Chicago, Illinois 60611-2795. Editor: John Webb, Librarian Emeritus, Washington State University Libraries, Pullman, WA 99164-5610. Annual subscription price, $55. Printed in U.S.A. with periodical-class postage paid at Chicago, Illinois, and other locations. As a nonprofit organization authorized to mail at special rates (DMM Section 424.12 only), the purpose, function, and nonprofit status for federal income tax purposes have not changed during the preceding twelve months.

EXTENT AND NATURE OF CIRCULATION

(Average figures denote the average number of copies printed each issue during the preceding twelve months; actual figures denote actual number of copies of single issue published nearest to filing date: June 2007 issue). Total number of copies printed: average, 5,354; actual, 5,280. Sales through dealers and carriers, street vendors, and counter sales: average, 0; actual 462. Paid or requested mail subscriptions: average, 4,283; actual, 4,193. Free distribution (total): average, 292; actual, 292. Total distribution: average, 5,028; actual, 4,947. Office use, leftover, unaccounted, spoiled after printing: average, 326; actual, 333. Total: average, 5,354; actual, 5,280. Percentage paid: average, 94.19; actual, 94.10.

Statement of Ownership, Management, and Circulation (PS Form 3526, September 2007) filed with the United States Post Office Postmaster in Chicago, October 1, 2007.