In this book, the term media is often treated in a very broad sense and refers to the nature of information perceived by humans. For example, auditive media include music, sound and voice. Visual media include text, graphics, still and moving pictures. These media concern mainly the human senses. Each medium defines presentation values (e.g., word, color, intensity) in presentation spaces (e.g., paper, computer monitor, speaker), which address our five senses [7].

Each presentation space has one or more presentation dimensions. A computer monitor has two space dimensions, while holography and stereophony need a third one. Time can occur as an additional dimension within each presentation space. Media are classified into two categories with regard to the time dimensions of their presentation space: discrete media and continuous media. Discrete media are composed of time-independent information items. Examples are text, graphics, and pictures. Continuous media refers to sound or motion video, where the presentation requires a continuous play-out in time. In other words, time-dependency between information items is part of the semantics of continuous media.

Media data is acquired through various sensors (e.g., camera, microphone) sensing the real-world, or synthesized using computers, or directly created by humans (e.g., news articles, blogs, tweets). Subsequently, media data may be subject to processing, compression, storage, delivery, analysis, retrieval, security protection, etc.

Some types of media (e.g., audio and video) and their applications typically exhibit the characteristics of high data rate, real-time constraint, and high power consumption, demanding significant resources such as storage, computing and bandwidth. Cloud computing and storage technologies that can pool resources and provide on-demand services become increasingly important for media applications. On the other hand, audio and video data are also loss-tolerant, and have prioritized components, which can be exploited in optimizing the performance of media systems.
Media is increasingly online nowadays thanks to the rapid evolution of the Internet and digital technologies. We have witnessed in recent years the proliferation of voice over IP (VoIP), over-the-top (OTT) video, IPTV, online news, blogs, tweets, facebook feed, Wikipedia. Data from different sources online is increasingly linked together, both through hyperlinks and semantically via linked data.

What is Multimedia then? The conventional definition is the use of a variety of communicative media (i.e., information intended for human consumption), including text, audio (sound and music), and visual (image, video, and graphics), and haptic data. Each type of media captures only partial information. Examples of multimedia include news videos, physiological data or social network posts. Multimedia can drastically enhance the storytelling prowess for any idea, news or campaign [1]. An example of this sort of interactive storytelling involving different pieces of multimedia is shown in Fig. 2.1.

Multimedia represents the emphasis on large varieties of data types. Today, the definition of multimedia has been extended to be much broader and more inclusive, covering a great variety of media, going far beyond the conventional types of audio and visual only [8]. There are new sensors capturing information in novel contexts of mobile, game, health, biomedical, environment, and many others. Multimedia represents the convergence of text, pictures, video, sound, and other forms of data into a single form. It is thus naturally richer with cross-modality and/or cross-domain features/data (e.g., tweets in Twitter and video on YouTube on the same topic). The power of multimedia and the Internet lies in the way in which information is linked.
The field of multimedia research studies computational algorithms, systems, and methods to handle multimedia data \[8\]. Multimedia research focuses on the scientific problems arising from the complementarity of different data sources. It studies how computer systems should be designed to process the different types of information available in an integrated fashion, including metadata (for example geo-tags) and the context in which the content is presented. Multimedia takes a systemic approach complementing other related fields such as signal processing, natural language processing, machine learning, software engineering, human-computer interaction, networking, by first learning the limitations of each individual source of information and then studying how the multiple sources can be used to provide a complete picture. In other words, multimedia unites signals from all our senses such as sight, sound, and touch, in an effort to generate a coherent perceptual experience as the human brain does.

2.1 Foundational Questions of Multimedia Research

Foundational questions of the multimedia research field include:

**Multimodal integration and synchronization**: By considering input from multiple sensors in an integrated fashion, multimedia problems can be solved more robustly and efficiently.

**Content analysis and retrieval**: Given diverse data sources in a very large amount of multimedia data, what is the best way to select data sources so that we can find a specific piece of information? Filling the semantic gap between low-level features (e.g., edges, color histogram) derived from the data and high-level concepts (e.g., people dancing) as perceived by human beings remains the most challenging problem.

**Context derivation**: What is the best way to derive context (the where, when, who, and what) for multimedia content analysis to minimize the semantic gap? For example, user interaction may be valid source of context to aid otherwise currently intractable content analysis problems. These may include user feedback in content search, or user actions (e.g., Like, Re-tweet) in online social network.

**Multimedia delivery**: The heterogeneity of existing communication networks and of content representations calls for the ability to interconnect such systems and formats, tackling source coding, network architecture, robustness, quality of service/experience, cross-layer optimization, security, and other problems. Using information regarding the user preferences and current state to provide a personalized service is critical for an effective system.

In today’s digital world and online communities, the usages of multimedia applications are ubiquitous. From image viewers on Facebook (social network) to filters on Instagram (social photos), from screen-casting on Twitch (video games) to 3D modeling in Maya (motion picture 3D), from music players like Spotify to video publishing sites like YouTube, and from e-book applications on Amazon’s Kindle to gif videos like Vine—multimedia applications govern human interaction
with machines. Several of these applications consist of challenging artificial intelligence problems. For example, in video publishing, the site needs to recommend relevant video to user based on what he/she is currently watching—a classic example of recommendation systems. In image search, the web site must parse a natural language text, extract semantics and retrieve related images. Moreover, query suggestion helps users to restructure their query based on available media content on the website and what other users’ have searched for. Brands use social media to quantify audience engagement, which requires intelligent analysis of user-generated media content to detect user profiles. These are all scenarios we shall discuss in this book, and describe how cross-domain data can help in improving individual multimedia applications.

Let us scrutinize these foundational questions in the light of media data generated in social networks. It is known to researchers that traditional image/video processing can imbue significant computational overhead, limiting its applicability in real time scenarios. Rather ironically, it is in real-time analysis of multimedia where most benefits are lie. Several recent works have found the multimodal integration of multimedia data from social networks can be used to improve prediction performance at no additional computational cost. These systems can thus scale to huge and complex data when deployed. Several traditional multimedia research problems are now being solved using additional social multimedia data.

Traditional methods are slower since they attempt to detect sensitive features without supporting meta-data. Meta-data is critical to understand the actual multimedia data. In fact, the more complex the data, the greater is the need of meta-data to comprehend it easily. The reason why social multimedia analysis often produces enhanced results than traditional approaches of multimedia analysis is the abundance of meta-data. Every activity on social network is directly or indirectly performed by humans and has inbuilt filters that allow users to themselves provide the system with meta-data. Thus, modern social networks like Facebook and Tumblr allow users to tag their media, ushering the age of data-generating products [2]. Data generating products like Facebook photo upload are designed in such a way so as to coerce users into supplying meta-data about the media. For example, a photo uploaded through phone can be tagged with location, landmark, time, face tags etc. The meta-data reduces ambiguity about objects in the image (object detection is a classic problem in image processing) producing an automatic channel for the multimedia. Since users are inclined not to mislead their friends or audience (YouTube), the tags they attach are mostly accurate. Thus, opportunities like search engine optimization is conventionally difficult to pull of with social multimedia data, because users have incentive not to game the system at the cost of deceiving their personal online social network. Content analysis and retrieval is easier with social multimedia data because the data is usually cleaner. Exceptions to this rule include Twitter, where the data tends to become noisy because of the short text format—leading to unconventional hashtags and bot followers.

At the same time, user generated data associated with media can provide tremendous context to it. Discussions on a Tumblr photo, re-tweets (RTs) on a
Twitter tweet and likes on a Facebook post not only provide semantic information about the media content, but also psychographic information related to the media. The latter is an invaluable resource. Before online social networks, it took Herculean resources and capricious survey tactics to retrieve psychographic information about some media. In present times, it only requires a researcher to write a customized crawler that accesses some social network API. Thus, both context of media and emotions about some context can be efficiently mined from social multimedia data.

Finally, the reach of media decides its acceptability or rejection. The delivery of content is as important as the content itself, which is why networks exists in most content industries (such as Cable TV) for the sole purposes of distribution. Social networks provide a competent way to observe, measure and analyze the diffusion of multimedia, its spread and its popularity. This is a critical commodity for content creators, distributors and advertisers since it allows you to quantify the quality or popularity of some content—based on factors like audience size. Social multimedia can be tracked much more comprehensively than traditional cable or media content, predominantly because there is little monopoly on the distribution architecture (Internet) and active end users who can interact with the service (unlike TV).

### 2.2 Angles into Social Multimedia Data

The fascinating thing about social multimedia is that since humans generate it, it contains a deep signal about the present human condition. The simplest definition of social multimedia is multimedia resources available by social media channels [3]. What this implies is there is a bunch of digital devices at our service, and when the multimedia content it generates can be seen or acted upon by others, it can be referred to as social multimedia. Thus, sources of multimedia content, which can be explicitly shared by users within a social, network or to cross-domain websites comprise social multimedia. Social multimedia content:

- promote communities, discussions and allow curation by users.
- can enhance traditional applications, such as business intelligence and advertising.
- can be employed to build novel applications such as story-telling or psychographic recommendations.

There are other multimedia that is shared outside social networks, but involve images, voice chats or videos. Examples of this are instant messengers or chat apps like WhatsApp (which was acquired by Facebook for $19 Billion). This is loosely termed as ‘dark social’, because a link to a webpage shared by a user on some of these apps cannot be tracked by target websites. Thus, the websites cannot recognize the origin of the traffic (unlike when the link is shared on Facebook directly). A final scenario is when social tools act upon media. This means
collaborative editing, for example in wikipedia. Although wikipedia is not explicit social network, editors can communicate on the platform, which generates unique patterns of association and activity [9]—thus multimedia on the website is collaboratively acted upon by humans.

There are three angles through which we can observe and study social multimedia data, namely (1) content format, (2) signal scope and (3) purpose.

**Content format:** Content format refers to the multimedia format of social multimedia. As shown in Fig. 2.2, there are several available formats in which media can be distributed, including audio, video, text, micro-text (non natural language or long-form), gifs and images. Examples of social multimedia data in music include spotify (the social jukebox) and 8 tracks (the socially curated radio) among others. Storify and slashdot helps create stories and aggregate articles with a social touch from across the web. They are usually in natural language. Both Instagram and Flickr allow users to follow others’ image stream, making them social image sharing media. Micromedia refers predominantly to tweets, as Twitter becomes ever more powerful in journalism (breaking news), brand media and communications. Giphy is a gif search engine, which collects gifs from the Internet and lets you share them. User-generated video content is the single biggest contributor of traffic on the web. Sites like YouTube and Vimeo are used by an
overwhelming majority of users to upload video content and watch videos. Twitch is a video platform for live-streaming events, where you can follow video streams and channels and users.

**Signal Scope:** The signal scope of social multimedia refers to the level at which data can be acted upon or perceived by users with respect to the social network (see Fig. 2.3). For example, we can like a Facebook message (of an individual) or we can click on trending topics in New York (for a group of people with geographical vicinity). The former is a micro-signal while the latter is a meso-signal. Thus, the scope refers to the base of the sample data size from which the signal is generated, whether an individual, a group or the entire network population. We shall go into further details about signal scope in the next chapter.

**Purpose:** The purpose of a social network is to let users communicate and discuss ideas, issues and content on the web. Recall that people want to do the same things that they do in real life, but only digitally because it is easier and faster. Thus, different social web domains allow people to focus on something they would like to do in real life, but achieve it through social means. For example, AirBnB allows you to treat a home owner as your travel lodging resource. Meetup lets users who share common interest engage in physical meetings in the neighborhood. Instapaper lets you bookmark web pages so you can read it offline on any device and share it with friends. Pinterest provides the same service, but in the form of pins instead of web pages. Foursquare lets users post their physical location. Yelp lets them find the best food around town with the help of others who have been to such restaurants. LinkedIn lets you network professionally. Twitch and Ustream allows users to life cast their actions, like singing or playing or just conversing. Notice the pattern here—social multimedia allows users to take some aspect of
their real life and achieve it digitally, be it networking for jobs or asking for restaurant suggestions. Figure 2.4 illustrates the different purposes of social multimedia.

### 2.3 Adoption of Social Media in Our Digital Lives

Users using these services are real, and they leave a data trail everywhere they go. Social multimedia data has dwarfed every other kind of traffic on the web. The statistics on the amount of social multimedia data generated is astonishing. Instagram has approximately 55 million posted daily. There are 8,500 likes per second. Twitter users send over 100,000 tweets per minute. YouTube has more than 1 billion unique users each month. 100 hours of video are uploaded every minute. Every minute, approximately 2,500 Foursquare check-ins are performed. Every 60 seconds on Facebook: 510 comments are posted, 293 thousand statuses are updated, and 136,000 photos are uploaded. Wondering what to do around your town? There are 315 thousand monthly meetups on ‘Meetup’, some could be in
your locality. We know that an exorbitant amount of traffic on the web is actually social multimedia traffic. This is entangling our digital lives with social multimedia. For example, YouTube reaches more US adults below 34 than any cable network. 1 million websites have been integrated into Facebook. 210,000 years of music have been played on Facebook. Even newsrooms are beginning to use tweets in order to add a voice of the people in addition to the experts in the studio.

There are several user interactions possible with social multimedia data. Although broadly confined by domain, some items are common such as likes/favorite some content, re-post/share/retweet some content or comment on some resource. All these activities add meta-data to the multimedia. This type of metadata (added by user actions) is very clean, less noisy and is preferred wherever possible in comparison to detection of visual features from image/video by computer vision algorithms. This is partly due to human labeling accuracy and partly due to lack of extra computational overhead. Therefore, social media data has become a disruptive platform for addressing many multimedia problems that could not be elegantly solved previously. For example, real-time social data is being utilized in semantic video indexing, image/video context annotation and visualizing political activity and flu outbreaks and emergency advisory systems. Social streams like Twitter are a good indicator of crowd sourcing activity of a social community. The information in social streams is real time, thus it can be used to learn about real life events quickly. Major world events in recent times, such as the Egyptian Revolution, the London Riots and the Japan Earthquake have been extensively captured using social streams such as Twitter and Facebook updates [4].

Social media has also largely affected existing models of communication and information retrieval. Akamai, a content-distribution company, recently reported that traffic from social sites has multiplied by five times in 2012, capping at 1 million requests per second. This has strong implications on traffic shaping for computer networks. Audiences are turning to social sites to ingest traditional news, e.g., 78% of web traffic to the New York Times website comes from just Facebook and Twitter combined. The rest 22% arrives from the organic web. Existing political and non-profit campaign prediction models, search tools and media recommendation have also changed to incorporate the massive amounts of social data generated every day.

One aspect of social micro-blogs like Twitter is its short text format, which is fast and real time. Thus, social media data hits the web faster than articles, images, or videos on the same topic. In the chain of digitization of a real-world event (Fig. 2.5), social stream data like tweets from Twitter are often the source of breaking news. In fact some famous breaking news in the past years has been captured first as tweets, including the death of Osama Bin Laden, the Hudson plane crash, announcement of the royal wedding etc. This property can be leveraged to resolve interesting real time applications, e.g. semantic video indexing [5] and topic evolution and topic tracking [6].

As mentioned earlier, a Social Multimedia Signal presumes human users as sensors and contains the spatio-temporal activity pattern of users (or user community) with respect to some multimedia content shared within the social network. For example, ‘Likes’ on Facebook, ‘Favorites’ on a YouTube video or a Retweet (RT) on Twitter indicates the popularity of media content. Such signals provide
perspective about a single media content, e.g., a shared video, a textual status message or an uploaded photo. There also exist social multimedia signals that provide contextual knowledge regarding the entire social ecosystem (as opposed to individual media content) within which media can be shared. Twitter trending topics is a fair reflection of the hot topics of discussion in the Twitter world.

Since a lot of this book is going to be about signals generated from social multimedia and their subsequent analysis, we must take a more thorough look at signal processing itself. The next chapter will explore the world of signals; time varying data streams and social multimedia signals in the light of signal processing.

References

Social Multimedia Signals
A Signal Processing Approach to Social Network Phenomena
Roy, S.D.; Zeng, W.
2015, X, 176 p. 95 illus., 79 illus. in color., Hardcover
ISBN: 978-3-319-09116-7