Toward Effective Movie Recommendations Based on Mise-en-Scène Film Styles

Yashar Deldjoo
Politecnico di Milan
Via Ponzio 34/5
20133, Milan, Italy
yashar.deldjoo@polimi.it

Mehdi Elahi
Politecnico di Milan
Via Ponzio 34/5
20133, Milan, Italy
mehdi.elahi@polimi.it

Massimo Quadrana
Politecnico di Milan
Via Ponzio 34/5
20133, Milan, Italy
massimo.quadrana@polimi.it

Paolo Cremonesi
Politecnico di Milan
Via Ponzio 34/5
20133, Milan, Italy
paolo.cremonesi@polimi.it

Franca Garzotto
Politecnico di Milan
Via Ponzio 34/5
20133, Milan, Italy
franca.garzotto@polimi.it

ABSTRACT

Recommender Systems (RSs) play an increasingly important role in video-on-demand web applications – such as YouTube and Netflix – characterized by a very large catalogs of videos and movies. Their goal is to filter information and to recommend to users only the videos that are likely of interest to them. Recommendations are traditionally generated on the basis of user's preferences on movies' attributes, such as genre, director, actors. Preferences on attributes are implicitly detected by analyzing the user's past opinions on movies.

However, we believe that the opinion of users on movies is better described in terms of the *mise-en-scène*, *i.e.*, the design aspects of a film production affecting aesthetic and style. Lighting, colors, background, and movements are all examples of mise-en-scène. Although viewers may not consciously notice film style, it still affects the viewer's experience of the film. The mise-en-scène highlights similarities in the narratives, as filmmakers typically relate the overall film style to reflect the story, and can be used to categorize movies at a finer level than with traditional film attributes, such as genre and cast. Two films may be from the same genre, but they can look different based on the film style.

In this paper, we present an ongoing work for the development of a novel application that offers a personalized way to search for interesting multimedia content. Instead of using the traditional classifications of movies based on attributes such as genre and cast, we use aesthetic movie features derived from film styles as determined by filmmaker professionals. Stylistic features of movies are extracted with an automatic video analysis tool and are used by our application to generate personalized recommendations and to help users in searching for interesting content.

CHItaly 2015, September 28-30, 2015, Rome, Italy

DOI: http://dx.doi.org/10.1145/2808435.2808460

CCS Concepts

•**Information systems** → *Recommender systems;*

Keywords

recommender systems, film style, content based, movie, video, mise-en-scène

1. INTRODUCTION

Nowadays, due to the rapidly growing of online video services, such as YouTube and Netflix, the information overload is getting more and more problematic and finding the right content is a challenging problem for the users [19]. *Recommender Systems* (RSs) are tools that tackle this problem by suggesting to users a set of videos that match their needs and interests [13, 20, 21, 1, 7].

The most popular approaches for RSs are based on *content-based* techniques [2, 14, 8], which suggest movies based on their explicit attributes. For instance, a content-based recommender system might recommends movies of the same genre or with the same actors as the movies that the user preferred in the past. The underlying idea of content-based recommender systems is that these attributes reflect the plot of the movie and the user's taste. One of the main drawbacks of content-based approaches is that the number of categories in which a movie can be classified is limited by the number of attributes. Moreover, in many application domains such as in YouTube, most of the video content available online is not tagged with attributes and, therefore, cannot be recommended.

Some authors observed that the perception of a movie in the eyes of a viewer is influenced by factors not related to genre and cast but to the *mise-en-scène*, *i.e.*, the style and design aspects of a film [5, 9]. Lighting, colors, background, costumes, camera movements, are all part of the mise-en-scène. As an example, according to the "classical hollywood narrative film style", if a character is walking across a room, the camera follows the character's movement. According to other film styles, the camera is fixed. Differently from traditional film attributes, such as genre and cast, which categorizes films based on the *plot* (i.e., the events that make up a story), the mise-en-scène highlights similarities in the *narrative structure* (i.e., how the events of a story are told), as filmmakers typically relate the overall film style to reflect the story. Two films may be from the same genre, but they can look different based on the film style. For example, "The Fifth Element"' and the "War of the Worlds"' are both sci-fi movies about an alien invasion. However, they are shot completely differently, with Luc Besson (The Fifth Element) using bright colors while Steven Spielberg (War of the Worlds) preferring dark scenes. Although a viewer may not consciously notice the two different film styles, they still affect the viewer's experience of the film. There are countless ways to create a film based on the same script simply through changing the mise-en-scène [9].

In this paper we describe an ongoing novel application that offers a personalized way to search for interesting movies through the analysis of film styles. Instead of using the traditional classifications of movies based on explicit attributes such as genre and cast, the application provides searching and recommendation capabilities based on aesthetic attributes derived from the low-level analysis of the video content. These aesthetic attributes are derived from film styles as determined by filmmaker professionals, and accurately match viewers' perceptions, as revealed by the first, ongoing studies. We would like to use this application to investigate our research hypothesis that styles–based recommender systems are able to provide better recommendations than content-based recommender systems. The application is composed of three main modules:

- a *low-level* feature extraction module which uses automatic video-analysis techniques to extract aesthetic features related to the style of a film, such as lighting, colors, movements;[1]
- a *style-based* recommender system which recommends to the user movies with a style similar to the style of the movies that the user liked in the past;
- an online application that can be used to search for interesting movies and to receive recommendations.[2]

The first two modules of the application have been completed and they now contain a catalog of 210 movies, 120 of which are uniformly sampled from 4 single genres (Action, Comedy, Drama, and Horror), and 90 from mixed genres. All the movies are characterized in terms of low-level style-based features. The development of the online application is still ongoing. Very preliminary results show that our style–based recommender system is able to provide better recommendations with respect to traditional content–based recommender systems. Once the online front-end is completed, we plan to launch an extensive empirical study to investigate the research hypothesis.

2. RELATED WORK

Content-based recommender systems build a profile of the user's preferences and interests by analyzing the features (attributes) of the products previously rated by the user. Recommendations are then generated by matching the features of the user profile with the features of other products in the catalog. In the media domain (i.e., music, movie), content-based recommendations are based on two different types of features: structured features related to the content (e.g., genre of the movie or music) or low-level features related to the intrinsic characteristics of the media content (e.g., brightness in video and pitch in music) [1, 7]. For example, in music recommendation many acoustic features, e.g. pitch, rhythm and timbre, can be extracted and used to find perceptual similar tracks [3, 4, 11, 18].

In this paper we focus on visual features, that can be extracted directly from the movie content itself and that can be correlated with the style of a movie. In video recommendation, a few works in the past have leveraged the visual features directly extracted from the visual content itself within the recommendation process. Scott et al. [17] describe an application that can be used to filter movies based on low-level features. Yang et al. [20] presented a video recommender system, VideoReach, which combines textual, visual and aural features to increase click-through-rate. Zhao et al. [21] propose a multi-task learning algorithm to integrate multiple ranking lists generated by exploring different information sources, visual content included.

However, none of these previous works has considered how visual features can effectively replace the other typical content information when they are not available. Moreover, none of these works have chosen the low-level features to be representative of different film styles.

3. METHOD DESCRIPTION

The first step in order to build a style-based video recommendation system is to find features that can bridge the gap between human-understandable semantics and aesthetic styles. These features must comply with human norms of perception and follow the style of the film - the rules directors of movies use to make a movie. By carefully studying the literature in the vision community, we selected a number of features that we think can best describe the main styles of a movie [16]. We conducted a features selected classification analysis in order to evaluate the usefulness of the features.

3.1 Visual features

A total of four categories of visual features was used in our experiment in order to represent each movie:

- *Average shot length*: A single camera action is known as a *shot* and the number of shots in a video can be indicative of the pace at which the movie is being created. For example, action movies usually contain rapid movements of the camera compared to dramas with many conversations between people in it. The corresponding average shot length in these genres is expected to be high (for actions) and low (for dramas).
- *Color variance*: There exist a strong level of correlation between variance of colors and the genre. Directors tend to adopt a large variety of bright colors for comedies and darker combination of colors for horrors. For each key frame represented in Luv color space, we compute the generalized color variance [16] as the representative of the amount of color variation in that key frame.

[1]The term *low-level* derives from the fact that style–based features are extracted from the movie through the low-level analysis of the video content, as opposed to traditional attributes which are manually assigned.

[2]The application will be accessible through the following link: http://recsys.deib.polimi.it/

- *Motion*: Motion in a video can be caused as the result of the camera movement (*i.e.* camera motion) or movements on part of the object being filmed (*i.e.* object motion). While the average shot length focuses on the former, it is desired that the latter type could be also captured accurately. For this purpose, a motion feature based on optical flow [10] was used to compute a robust estimate of the pixel velocities over a sequence of images being filmed.

- *Lightening*: Lightening is another discriminating factor between movie genres and can be used as a key playing factor to control the type of emotion desired to be induced to a user. For example, comedies often contain abundance of light with a low *key-to-fill* ratio, *i.e.*, a low ratio between the brightest and dimmest light. This concept in cinematography is known as *high-key lightening*. On the other hand, horror or noir films use low-key lightening, *i.e.* low amount of light and a high key-to-fill ratio.

3.2 Recommendation algorithm

Videos can be recommended to users by choosing movies that have similar characteristics to the ones previously seen by the user. This approach is usually referred to as *content-based recommendation*, since only the content features of the videos are exploited to generate a personalized list of suggested movies to the user. A widely adopted state-of-the-art algorithm in this family is called *k-nearest-neighbors* content-based recommender. In this algorithm the preference of a user for an unseen movie, called the *target* movie, is modelled as the weighted average of her preferences toward previously seen movies, called the *known* movies. Each preference score is weighted proportionally to the similarity between the target and known movie this score refers to. Consequently, all the unseen movies in the catalog can be scored according to their predicted preference, and a list of the top-n scored movies is presented to the user. There exist many similarity metrics that can be plugged into the recommender algorithm, e.g. *Pearson correlation* or *cosine similarity*, and each one will lead to different recommendation lists. In this work we adopted k-nearest-neighbors with cosine similarity as the algorithm to generate personalized lists of suggested movies to users.

4. USER STUDY

We have developed a demo application, that is going to be used in a live user study, we planned to conduct. Figure 1 llustrates sample screenshots of the demo. First a user is registered by entering her basic demographics. Then, she is asked to clarify whether or not she is an expert in cinema or movie making domain (top screenshot). This is done in order to later group the users according to their expertise in the domain. Later, the user can enter her preferences in terms of genres (top screenshot) and the visual low-level features she might be interested to enter (middle screenshot). Finally, the user is presented a list of recommendations generated according to her preferences (bottom screenshot). These recommendations are based on either content-based recommendations according to genre (method 1), or Style-based recommendations according to the aesthetic visual features (method 2). As a relevance feedback to the recommendations, the user is asked to check the relevancy of the movies and to give her ratings to them. These ratings are later used to compare the quality of the two recommendation methods.

We also plan to provide to user a questionnaire that measures the quality of the recommendations and user satisfaction, as well as the usability of the system. For the first metric, we adopt one of the well-known questionnaires [12, 15], which provides a standard questionnaire for perceived recommendation quality and choice satisfaction. For the second metric, we adopt SUS (System Usability Scale)[6] that has become a standard for perceived system usability analysis.

5. CONCLUSION AND FUTURE WORK

In this paper we describe an ongoing research which its goal is to develop an application that offers personalized way to search for interesting movies through the analysis of film styles. Our application will provide not only the searching capability, but also will benefit from a *Style-based* recommendation engine. Hence, it exploits aesthetic attributes derived from the low-level analysis of the video content, rather than the traditional attributes of movies, *i.e.*, explicit attributes such as genre and cast.

We have developed a demo application, that we are using it to test our research hypothesis, *i.e.*, styles-based video recommender systems are able to provide better recommendations than classical content-based video recommender systems. Our preliminary result, obtained from first ongoing studies, has shown promising quality for our recommender system in comparison to the classical content-based system.

6. REFERENCES

[1] J.-w. Ahn, P. Brusilovsky, J. Grady, D. He, and S. Y. Syn. Open user profiles for adaptive news systems: help or harm? In *Proceedings of the 16th international conference on World Wide Web*, pages 11–20. ACM, 2007.

[2] M. Balabanović and Y. Shoham. Fab: Content-based, collaborative recommendation. *Commun. ACM*, 40(3):66–72, Mar. 1997.

[3] D. Bogdanov and P. Herrera. How much metadata do we need in music recommendation? a subjective evaluation using preference sets. In *ISMIR*, pages 97–102, 2011.

[4] D. Bogdanov, J. Serrà, N. Wack, P. Herrera, and X. Serra. Unifying low-level and high-level music similarity measures. *Multimedia, IEEE Transactions on*, 13(4):687–701, 2011.

[5] D. Bordwell and K. Thompson. *Film art: an introduction.* McGraw Hill, 2008.

[6] J. Brooke. Sus: A quick and dirty usability scale. *Usability evaluation in industry*, 189:194, 1996.

[7] I. Cantador, M. Szomszor, H. Alani, M. Fernández, and P. Castells. Enriching ontological user profiles with tagging history for multi-domain recommendations. 2008.

[8] P. Cremonesi, F. Garzotto, S. Negro, A. Papadopoulos, and R. Turrin. Comparative evaluation of recommender system quality. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 1927–1932, New York, NY, USA, 2011. ACM.

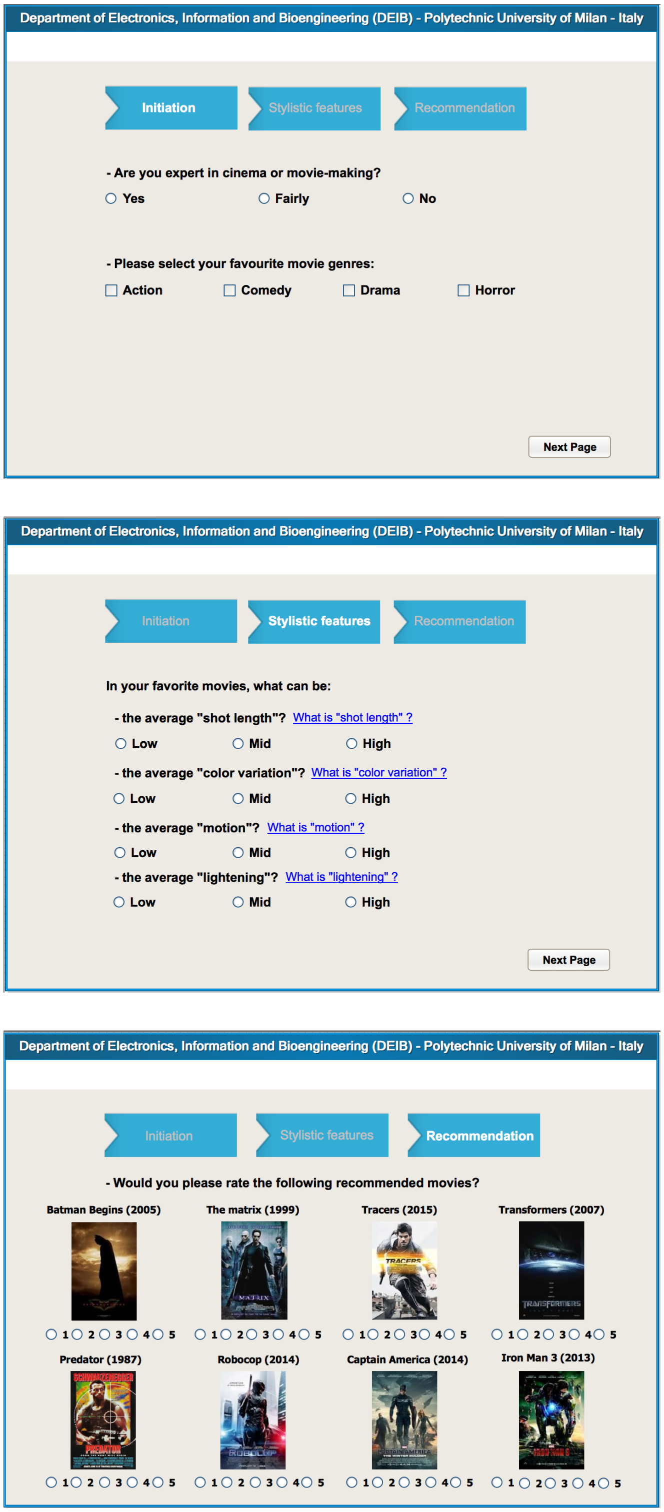

Figure 1: Sample screenshots of the application: Initiation (top), Preference elicitation for stylistic features (middle), and Recommendation (bottom)

[9] J. Gibbs. *Mise-en-scène: Film Style and Interpretation*. Short cuts : introduction to film studies. Wallflower, 2002.

[10] B. K. Horn and B. G. Schunck. Determining optical flow. In *1981 Technical Symposium East*, pages 319–331. International Society for Optics and Photonics, 1981.

[11] P. Knees, T. Pohle, M. Schedl, and G. Widmer. A music search engine built upon audio-based and web-based similarity measures. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 447–454. ACM, 2007.

[12] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, and C. Newell. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):441–504, 2012.

[13] T. Lehinevych, N. Kokkinis-Ntrenis, G. Siantikos, A. S. Dogruöz, T. Giannakopoulos, and S. Konstantopoulos. Discovering similarities for content-based recommendation and browsing in multimedia collections.

[14] M. J. Pazzani and D. Billsus. The adaptive web. chapter Content-based Recommendation Systems, pages 325–341. Springer-Verlag, Berlin, Heidelberg, 2007.

[15] P. Pu, L. Chen, and R. Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 157–164. ACM, 2011.

[16] Z. Rasheed, Y. Sheikh, and M. Shah. On the use of computable features for film classification. *Circuits and Systems for Video Technology, IEEE Transactions on*, 15(1):52–64, 2005.

[17] D. Scott, F. Hopfgartner, J. Guo, and C. Gurrin. Evaluating novice and expert users on handheld video retrieval systems. In *Advances in Multimedia Modeling*, pages 69–78. Springer, 2013.

[18] K. Seyerlehner, M. Schedl, T. Pohle, and P. Knees. Using block-level features for genre classification, tag classification and music similarity estimation. *Submission to Audio Music Similarity and Retrieval Task of MIREX 2010*, 2010.

[19] J. Shen, M. Wang, S. Yan, and P. Cui. Multimedia recommendation. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1535–1536. ACM, 2012.

[20] B. Yang, T. Mei, X.-S. Hua, L. Yang, S.-Q. Yang, and M. Li. Online video recommendation based on multimodal fusion and relevance feedback. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 73–80. ACM, 2007.

[21] X. Zhao, G. Li, M. Wang, J. Yuan, Z.-J. Zha, Z. Li, and T.-S. Chua. Integrating rich information for video recommendation with multi-task rank aggregation. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1521–1524. ACM, 2011.