Quantifying Conflicts in Narrative Multimedia by Analyzing Visual Storytelling Techniques*

O-Joun Lee[1,[0000−0001−8921−5443]], Jin-Taek Kim[2], and Eun-Soon You[3,**]

[1] Department of Artificial Intelligence, The Catholic University of Korea
43, Jibong-ro, Bucheon-si, Gyeonggi-do 14662, Republic of Korea
ojlee@catholic.ac.kr
[2] Future IT Innovation Laboratory, Pohang University of Science and Technology
77, Cheongam-ro, Nam-gu, Pohang-si, Gyeongsangbuk-do 37673, Republic of Korea
jintaek@postech.ac.kr
[3] Department of French Language and Culture, Inha University
100, Inha-ro, Michuhol-gu, Incheon 22201, Republic of Korea
jiwony71@gmail.com

Abstract. This study aims at measuring conflict degrees of each shot in visual narrative multimedia (e.g., movies and TV series) by analyzing visual storytelling techniques, such as camerawork. To describe incidents in stories, directors use the techniques as like as visual language. Thus, visual storytelling techniques used in a shot should be correlated with incidents shown by the shot. In this study, we first present various taxonomies of the visual storytelling techniques and discuss which techniques have more correlations with conflicts than the others. Then, based on usages of the techniques in each shot, we measure intensity of conflicts described by the shot. Finally, we validated correlations of visual storytelling techniques with stories' content by examining correlations of the proposed conflict measurement with conflict degrees annotated by scholars and practitioners in film studies.

Keywords: Computational Narrative Understanding, Camerawork Analysis, Conflict Measurement, Visual Storytelling

1 Introduction

Conflicts are a significant feature of the narrative analysis since stories are led by conflicts around their protagonists [14,15,25,21]. For example, if we can compare shots in terms of their conflict degrees, highlight clips of movies can be composed by gathering top-N shots according to the conflict degrees. Existing studies for measuring conflicts employed mainly two approaches: (i) character network (i.e., social networks of

[*] Copyright © by the paper's authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). In: N. D. Vo, O.-J. Lee, K.-H. N. Bui, H. G. Lim, H.-J. Jeon, P.-M. Nguyen, B. Q. Tuyen, J.-T. Kim, J. J. Jung, T. A. Vo (eds.): Proceedings of the 2nd International Conference on Human-centered Artificial Intelligence (Computing4Human 2021), Da Nang, Viet Nam, 28-October-2021, published at http://ceur-ws.org

[**] Corresponding author.

fictional characters) analysis [9,12,2,7] and (ii) sentiment analysis [9,6]. The character network-based methods assume that conflicts in stories accompany frequent interactions between characters, which cause changes in structures of character networks [7,10]. Thus, these methods quantify conflicts by measuring structural changes in character networks. However, they cannot consider meanings of individual interactions/incidents. Sentiment analysis-based methods resolve this limitation. They apply the sentiment analysis on emotional words in dialogues or facial/vocal expressions of actors/actresses. This approach supposes that conflicts accompany intense and negative emotions. However, advents of new media (e.g., webtoons and webnovels) hinder applying one sentiment analysis tool on the entire narrative multimedia corpus, even if the tool can analyze context and figurative expressions [9].

Beyond the two approaches, a few studies [17,1] focused on characteristics of visual storytelling, such as camerawork. In visual narrative multimedia (e.g., movies and TV series), the camerawork is a significant channel of storytelling as much as dialogue and acting [4,18]. Canini et al. [3] classified shots according to shot sizes (i.e., distances between cameras and subjects). Wang and Cheong [26] have proposed a shot taxonomy based on camera motions and shot sizes and classified shots according to their taxonomy. Rasheed et al. [19] classified movies into genres by analyzing shot lengths and color usages in shots. Svanera et al. [24] attempted to recognize movies' directors by analyzing shot sizes and lengths. Despite these various attempts, the existing studies did not consider theoretical models and practices for camerawork in the film studies. They merely supposed that physical features of shots have narrative meanings. Although Svanera et al. [23] picked over the shoulder (OTS) shots as a shot type correlated with tensions and have proposed a method for detecting OTS shots, OTS shots are only one of various shot types for describing tensions.

In film studies, there have already been various shot taxonomies based on camerawork, and uses of each shot type have also been widely studied [16,20]. Therefore, if we know shot types and their meanings, we can analyze shots' content by detecting usages of camerawork in the shots. This study first introduces shot taxonomies and criteria of the taxonomies by focusing on shot types related to conflicts. Then, we propose a conflict measurement based on usages of camerawork. Finally, we validate whether the camerawork has correlations with shots' content by examining accuracy of the proposed measurement.

2 Conflict Measurement based on Camerawork

A long history of visual narrative multimedia makes directors follow formulaic grammars of visual storytelling. Directors are aware of effective camerawork to deliver incidents to audiences enclosing their intentions. The camerawork includes various features, such as shot sizes, camera angles, and screen composition, and these features are criteria of shot taxonomies in film studies. Thus, we discuss correlations of the features and shot types with describing conflicts and quantify conflicts based on the shot types.

(a) (b)

Fig. 1. Shot types according to the number of characters. (**a**) A two shot in ‘Stoker’ (2013). (**b**) A group shot in ‘The Godfather’ (1972).

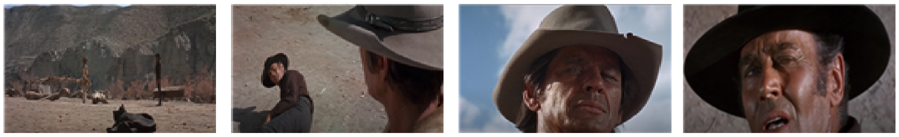
2.1 Correlations of Shot Types with Conflicts

Number of Characters Shots are categorized into ‘one shots,’ ‘two shots,’ and ‘group shots’ according to the number of characters in the shots [16]. For depicting conflicts, two shots clarify significance of relationships between two characters. Also, by employing other storytelling techniques together, we can set meanings of the relationships [20]. For example, portions of characters’ faces on frames can imply their power relationship. On the other hand, since group shots should use smaller shot sizes than two shots [16], they have difficulties for describing individual relationships of characters.

Fig. 1 (**a**) presents conversation between ‘Charles’ and ‘India’ in ‘Stoker’ (2013). ‘Charles’ takes a larger area (shorter camera distance) than ‘India,’ and it makes audiences aware of importance of the conversation and dominance of ‘Charles.’ (**b**) shows a mafia’s meeting (group shot) in ‘The Godfather’ (1972). From the shot, it is not easy to recognize relationships of individual characters.

Screen Composition Screen composition indicates how entities on frames (e.g., actors/actresses, scenery, and props) are located. Using the screen composition, we can subdivide the two shots. Fig. 2 show three ‘two shots’ in ‘Once upon a time in the west’ (1968), but (**b**) and (**c**) make variations using unique screen compositions. (**b**) is an OTS shot that shows a character over the shoulder of another character [16]. Thus, OTS shots hide facial expressions or behaviors of one side [20]. This composition describes characters’ relationships more intimately or their conflicts more intensely than normal two shots [23]. (**c**) is a shot reverse shot, which shows a character looking at another character (often off-screen) and then shows the latter character looking at the former one [22]. By showing two characters alternately, this shot type describes emotional reactions of the characters for each other’s behaviors. Thus, shot reverse shots are frequently used in climaxes of conflicts.

Directions of Characters’ Eyes Eye directions of characters are a kind of ‘charade’ (i.e., nonverbal storytelling), such as facial expressions and gestures [5]. Among the eye



(a) (b) (c)

Fig. 2. Shot types related to the screen composition. (**a**) to (**c**) A two shot, an OTS shot, and a shot reverse shot in ‘Once upon a time in the west’ (1968).

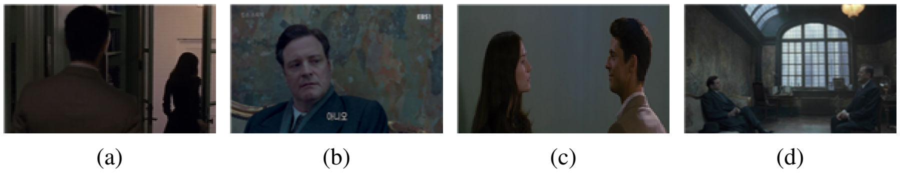


Fig. 3. Examples of the eye direction. (**a**) and (**b**) Eye aversion shots in 'Stoker' (2013) and 'King's Speech' (2010). (**c**) and (**d**) Eye contact shots in the two movies.

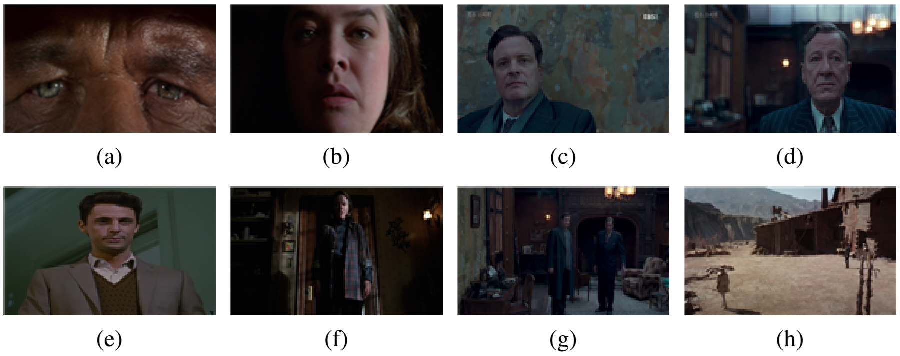


Fig. 4. Examples of shot sizes. (**a**) An extreme close up shot in 'Once upon a time in the west' (1968). (**b**) A close up shot in 'Misery' (1990). (**c**) and (**d**) Medium close up shots in 'King's Speech' (2010). (**e**) A medium shot in 'Stoker' (2013). (**f**) A medium long shot in 'Misery' (1990). (**g**) A long shot in 'King's Speech' (2010). (**h**) An extreme long shot in 'Once upon a time in the west' (1968).

directions, eye aversion and eye contact symbolize power relationships between characters. Eye contact is correlated with tensions and confrontations between characters. Fig. 3 (**d**) shows eye contact between 'King George VI' and 'Lionel' in 'King's Speech' (2010), while sitting apart. This shot describes conflicts between the two characters for appellations and speech therapy. (**c**) depicts the first conversation between 'Charles' and 'India' in 'Stoker' (2013). Different from Fig. 1 (**a**), their eye contact shows equal relationships among them and raises tensions. Contrarily, eye aversion implies conflicts that one side is passive. In Fig. 3 (**a**), eyes of 'Charles' follow 'India' obstinately, but 'India' avoids. On lots of shots in 'King's Speech' (2010), including (**b**), 'King George VI' avoids eyes of 'Lionel' when 'Lionel' asks uncomfortable questions.

Shot Size Shot sizes indicate relative sizes of subjects (e.g., characters) on frames. We classify shots into seven types according to shot sizes: extreme close up shots, close up shots, medium close up shots, medium shots, medium long shots, long shots, and extreme long shots (from big to small shot sizes) [16]. Bigger shot sizes have more correlations with conflicts, since they are effective to describe characters' emotions using facial expressions, as shown in Fig. 4 (**a**) and (**b**) [16,20]. Medium shot sizes can show body language, facial expressions, and backgrounds altogether, as shown in (**c**) to (**f**) [16,20]. Directors use them to explain incidents, narrative worlds, or characters' motivations. The other small shot sizes aim at describing spatial backgrounds [16]. Also,

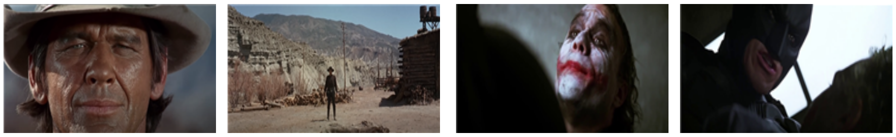


Fig. 5. Cross-cutting contrary shots. (**a**) Contrary shot sizes in 'Once upon a time in the west' (1968). (**b**) Contrary camera angles in 'The dark knight' (2008).

contrasts of shot sizes are more effective in escalating tensions than they are solely used. In Fig. 5 (**a**) from 'Once upon a time in the west' (1968), the close up shot (left) describes tensions of 'Harmonica,' and the extreme long shot (right) shows his enemy from his viewpoint.

Camera Angle Camera angles indicate angles between cameras and subjects and correspond to audiences' viewpoints. Among shot types according to camera angles, high and low angle shots have more correlations with conflicts than the others [13]. High angle shots are taken from higher locations than eye-levels [5]. Since audiences look down on characters (or other subjects), these shots show overall situations and describe the characters as weak and fragile ones [20]. On the other hand, low angle shots make audiences look up characters and give authorities and powers to the characters [5,20]. Cross-cutting high and low angle shots intensifies conflicts. In Fig. 5 (**b**), 'The dark knight' (2008) presents the high angle shot (left) that show 'Joker' and the low angle shot (right) for 'Batman,' alternately, to contrast positions of 'Joker' with 'Batman.'

2.2 Quantitative Measurements of Conflict Degrees

Our conflict measurement focuses on usages and combinations of the camerawork techniques discussed in the previous section. Although we do not deal with methods for detecting camerawork in shots, it requires only simple computer vision techniques [1], and screenplays mostly include annotations for camerawork [8]. First, two shots have tighter correlations with conflicts than the others, and the two variations of two shots tighten the correlations. For the i-th shot (s_i), we quantify its conflict degree as:

$$C_N(s_i) = I_T(s_i) \times w_T + I_O(s_i) \times w_O + I_R(s_i) \times w_R, \quad w_T < w_O, w_R \tag{1}$$

where $I_T(s_i)$, $I_O(s_i)$, and $I_R(s_i)$ are indicator functions for two shots, OTS shots, and shot reverse shots, respectively, and w_T, w_O, and w_R are weighting factors for the three shot types. As a preliminary study, we set w_T, w_O, and w_R as 0.5, 1.0, and 1.0, respectively.

Both eye aversions and eye contacts describe conflicts between two characters. However, eye aversions usually depict unexposed conflicts, while eye contacts show that conflicts finally boil over. Conflicts expressed by eye directions can be measured as:

$$C_E(s_i) = I_A(s_i) \times w_A + I_G(s_i) \times w_G, \quad w_T < w_A < w_G, \tag{2}$$

where $I_A(s_i)$ and $I_G(s_i)$ are indicator functions for eye aversion shots and eye contact shots, respectively, and w_A and w_G refer to weighting factors for the two shot types. We set w_A and w_G as 0.7 and 1.0, respectively.

Bigger shot sizes are more effective to describe conflicts. We set conflict degrees of the seven shot types with regular intervals as: 1, $\frac{5}{6}$, $\frac{4}{6}$, $\frac{3}{6}$, $\frac{2}{6}$, $\frac{1}{6}$, and 0. However, combinations of contrary shot sizes emphasize conflicts. Thus, we check shot sizes of adjacent shots within each scene, since scenes are narrative units describing independent incidents [14]. When s_j is a consequent shot of s_i, conflict degrees according to shot sizes, $C_D(s_i)$, can be updated as:

$$C_D(s_i) := C_D(s_i) + I_C(s_i, s_j) \times |C_D(s_i) - C_D(s_j)|, \tag{3}$$
$$I_C(s_i, s_j) = \begin{cases} 1, & \text{if } |C_D(s_i) - C_D(s_j)| \geq \frac{2}{6}, \\ 0, & \text{otherwise.} \end{cases},$$

where $I_C(s_i, s_j)$ is an indicator function for cross-cutting of contrary shot sizes.

For camera angles, we can consider three cases: high angle, low angle, and combinations of high and low angles. It is difficult to say which one is correlated with more intense conflicts among triumphs and despairs of characters. However, contrasts between the two emotions have higher correlations to conflicts than the monotonous ones. Thus, we first set conflict degrees in both high and low angle shots as 1.0 and update the degrees according to usages of cross-cutting. When s_i and s_j are adjacent in a scene, conflict degrees according to camera angles, $C_A(s_i)$, can be updated as:

$$C_A(s_i) := C_A(s_i) + I_{HL}(s_i, s_j) \times w_{HL}, \tag{4}$$

where $I_{HL}(s_i, s_j)$ refers to an indicator function for whether s_i and s_j are taken by contrary camera angles, and w_{HL} is a weighting factor for the cross-cutting of contrary camera angles. We set w_{HL} as 1.

To quantify conflicts in shots, we aggregate the four proposed measurements using the arithmetic mean: $C(s_i) = \frac{1}{4} \times [C_N(s_i) + C_E(s_i) + C_D(s_i) + C_A(s_i)]$. Furthermore, visual narrative multimedia consist of various units on multiple granularity levels (e.g., shots $\in$ scenes $\in$ sequences $\in$ acts $\in$ movies) [11]. We can measure a conflict degree of a coarser unit by aggregating conflict degrees of shots included in the unit.

3 Evaluation

We evaluated the proposed measurements and validated correlations of the visual storytelling techniques with conflicts by comparing the proposed measurements with conflict degrees felt by human evaluators. To secure objectivity of the human evaluation, we have two options: composing a large-scale evaluator group or a reliable expert group. Since quantifying conflict degrees in each shot with consistent criteria is not very easy for general audiences, we composed an expert group that consists of five scholars and practitioners in film studies[4].

First, the evaluators annotated camerawork used in each shot according to the five criteria presented in Sect. 2.1. Then, they also annotated conflict degrees in each shot with integers from 0 to 5. We compared the manually annotated conflict degrees with the

[4] We would like to express thanks to our evaluator group, Dr. Choi, Inkyung, Mr. Heo, Sung Phil, Ms. Han, Jeongmin, Ms. Kim, Hayeong, and Ms. Kwak, Bo Eun.

Table 1. Experimental results for the four proposed measurements (C_N, C_E, C_D, and C_A in Sect. 2.2) and the combination of the four measurements (C).

	C	C_N	C_E	C_D	C_A
Average	0.74	0.41	0.64	0.67	0.61
S.D.	0.26	0.26	0.21	0.19	0.22

proposed measurements calculated using the camerawork annotations. The comparison was conducted by the Pearson correlation coefficient. If coefficients are close to 1, the proposed measurements are accurate, and our hypothesis is evident. We calculated PCC for each evaluator and averaged them for each experimental subject. We chose five movies as experimental subjects: 'King's Speech' (2010), 'Stoker' (2013), 'Once Upon a Time in the West' (1968), 'Misery' (1990), and 'The Dark Knight' (2008). Due to the manual annotation, this experiment has a limited scale. Thus, we attempted to choose representative movies of various genres. Table 1 presents experimental results.

The proposed measurements exhibited reasonable accuracy in terms of both accuracy and variance. Especially, the combination of the four measurements outperformed cases that the four measurements are independently used. This point underpins that combinations of camerawork make synergy effects as we expected. However, C_N exhibited significantly low accuracy than the other measurements. Also, a combination of the other three measurements outperformed the combination of all the measurements (accuracy: 0.74 and variance: 0.23). We should reconsider correlations of conflict degrees with the number of characters and screen composition. Among the remaining measurements, C_D (shot sizes) exhibited the highest accuracy and the lowest variance. When we did not consider cross-cutting of contrary shot sizes, the shot size exhibited much lower performance than C_D (accuracy: 0.58 and variance: 0.26). Similarly, when cross-cutting of contrary camera angles was not reflected, the camera angle underperformed C_A (accuracy: 0.56 and variance: 0.25). This result validates our assumption that the cross-cutting of contrary shots emphasizes conflicts. Conclusively, correlations of visual storytelling techniques with story content were validated by the reasonable accuracy of the proposed measurements.

4 Conclusion

We proposed the measurements for conflict degrees in visual narrative multimedia based on usages of camerawork. Despite the reasonable accuracy of the proposed measurement, our experiment had limitations on its scale. Also, conflict-related shots in Sect. 2.1 are only a part of visual storytelling techniques. Our further research will be focused on extending and enriching our dataset.

References

1. Bak, H.Y., Park, S.B.: Comparative study of movie shot classification based on semantic segmentation. Applied Sciences **10**(10), 3390 (May 2020). https://doi.org/10.3390/app10103390

2. Bost, X., Gueye, S., Labatut, V., Larson, M., Linarès, G., Malinas, D., Roth, R.: Remembering winter was coming. Multimedia Tools and Applications **78**(24), 35373–35399 (Sep 2019). https://doi.org/10.1007/s11042-019-07969-4
3. Canini, L., Benini, S., Leonardi, R.: Classifying cinematographic shot types. Multimedia Tools and Applications **62**(1), 51–73 (Nov 2011). https://doi.org/10.1007/s11042-011-0916-9
4. Doane, M.A.: The close-up: Scale and detail in the cinema. Differences: A Journal of Feminist Cultural Studies **14**(3), 89–111 (Jan 2003). https://doi.org/10.1215/10407391-14-3-89
5. Hayward, S.: Cinema Studies: The Key Concepts. Routledge Key Guides, Routledge, Abingdon-on-Thames, United Kingdom, 4th edn. (Feb 2013)
6. Jung, J.J., You, E., Park, S.: Emotion-based character clustering for managing story-based contents: a cinemetric analysis. Multimedia Tools and Applications **65**(1), 29–45 (jun 2013). https://doi.org/10.1007/s11042-012-1133-x
7. Lee, O.J., Jung, J.J.: Character network embedding-based plot structure discovery in narrative multimedia. In: Akerkar, R., Jung, J.J. (eds.) Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics (WIMS 2019). pp. 15:1–15:9. ACM, Seoul, Republic of Korea (Jun 2019). https://doi.org/10.1145/3326467.3326485
8. Lee, O.J., Jung, J.J.: Integrating character networks for extracting narratives from multimodal data. Information Processing and Management **56**(5), 1894–1923 (Sep 2019). https://doi.org/10.1016/j.ipm.2019.02.005
9. Lee, O.J., Jung, J.J.: Modeling affective character network for story analytics. Future Generation Computer Systems **92**, 458–478 (Mar 2019). https://doi.org/10.1016/j.future.2018.01.030
10. Lee, O.J., Jung, J.J.: Story embedding: Learning distributed representations of stories based on character networks. Artificial Intelligence **281**, 103235 (Apr 2020). https://doi.org/10.1016/j.artint.2020.103235
11. Lee, O.J., Jung, J.J., Kim, J.T.: Learning hierarchical representations of stories by using multi-layered structures in narrative multimedia. Sensors **20**(7), 1978 (Apr 2020). https://doi.org/10.3390/s20071978
12. Liu, C., Last, M., Shmilovici, A.: Identifying turning points in animated cartoons. Expert Systems with Applications **123**, 246–255 (Jun 2019). https://doi.org/10.1016/j.eswa.2019.01.003
13. McCain, T.A., Chilberg, J., Wakshlag, J.: The effect of camera angle on source credibility and attraction. Journal of Broadcasting **21**(1), 35–46 (Jan 1977). https://doi.org/10.1080/08838157709363815
14. McKee, R.: Story: Substance, Structure, Style and the Principles of Screenwriting. HarperCollins, New York, NY, USA (Nov 1997)
15. McKee, R.: Dialogue: The Art of Verbal Action for Page, Stage, and Screen. Twelve, New York City, NY, USA (Jul 2016)
16. Mercado, G.: The Filmmaker's Eye: Learning (and Breaking) the Rules of Cinematic Composition. Taylor & Francis Ltd., Oxfordshire, United Kingdom, 3rd edn. (Sep 2010), `https://www.ebook.de/de/product/11443372/gustavo_mercado_the_filmmaker_s_eye.html`
17. Qu, W., Zhang, Y., Wang, D., Feng, S., Yu, G.: Semantic movie summarization based on string of ie-rolenets. Computational Visual Media **1**(2), 129–141 (Jun 2015). https://doi.org/10.1007/s41095-015-0015-3
18. Quigley, P.: Eisenstein, montage, and 'filmic writing'. In: Antoine-Dunne, J., Quigley, P. (eds.) The Montage Principle, Critical Studies, vol. 21, pp. 153–169. Brill Rodopi, Leiden, Netherlands (Sep 2004)

19. Rasheed, Z., Sheikh, Y., Shah, M.: On the use of computable features for film classification. IEEE Transactions on Circuits and Systems for Video Technology **15**(1), 52–64 (Jan 2005). https://doi.org/10.1109/tcsvt.2004.839993
20. Sijll, J.V.: Cinematic Storytelling. Publishers Group UK, London, United Kingdom, 2nd edn. (Aug 2007), `https://www.ebook.de/de/product/4219016/jennifer_van_sijll_cinematic_storytelling.html`
21. Sikov, E.: Film Studies, second edition. Film and Culture Series, Columbia University Press, New York City, NY, USA, 2nd edn. (Jun 2020), `https://www.ebook.de/de/product/39399746/ed_sikov_film_studies_second_edition.html`
22. Spadoni, R.: The figure seen from the rear, vitagraph, and the development of shot/reverse shot. Film History **11**(3), 319–341 (1999)
23. Svanera, M., Benini, S., Adami, N., Leonardi, R., Kovács, A.B.: Over-the-shoulder shot detection in art films. In: Proceedings of the 13th International Workshop on Content-Based Multimedia Indexing (CBMI 2015). IEEE, Prague, Czech Republic (Jun 2015). https://doi.org/10.1109/cbmi.2015.7153627
24. Svanera, M., Savardi, M., Signoroni, A., Kovacs, A.B., Benini, S.: Who is the film's director? authorship recognition based on shot features. IEEE MultiMedia **26**(4), 43–54 (Oct 2019). https://doi.org/10.1109/mmul.2019.2940004
25. Truby, J.: The anatomy of story: 22 steps to becoming a master storyteller. Farrar, Straus and Giroux, New York City, NY, USA (Oct 2008)
26. Wang, H.L., Cheong, L.F.: Taxonomy of directing semantics for film shot classification. IEEE Transactions on Circuits and Systems for Video Technology **19**(10), 1529–1542 (Oct 2009). https://doi.org/10.1109/tcsvt.2009.2022705