Who is the film's director? Authorship recognition based on shot features

Michele Svanera
University of Brescia

Mattia Savardi
University of Brescia

Alberto Signoroni
University of Brescia

András Bálint Kovács
ELTE University

Sergio Benini
University of Brescia

We show how to perform automatic attribution of movie's authorship by the statistical analysis of two simple formal features: the length of camera takes (shot duration), and the distance between the camera and the subject (shot scale). Experiments include 143 films by 8 distinguishable directors over 70 years of historiography of author cinema.

The impression that film authors have very different and recognizable styles is very common: the omnipresence of one-point perspective scenes in Kubrick's production, Almodovar's obsession for red objects, Tarantino's shots taken from inside car trunks, etc..

However, looking for features which are distinctive of a single director, and remain invariants in all author's movies, is a hard task. Films from the same director can be in fact very different from both technical and stylistic point of view, taking into account that an author's filmography may span for more than half a century (such as Bergman's). Other reasons of stylistic changes within the production of a single director are, for example, the innovation in individual style, or differences in narrative or genre: think for example how different is Scorsese's *Hugo* from 2011 - an historical adventure drama film entirely shot in 3D – from his other work *The Wolf of Wall Street* - a biographical black comedy crime film - released just two years later.

Due to these variations, a fascinating research question is whether it is possible to automatically isolate in movies those elements which are distinctive of a single author and remain constant through time and styles. Aiming to characterize these individual characteristics of a director, we should first analyze those formal features which appear both conspicuously and frequently, and are under her/his direct control,[1] such as: *shot duration*, i.e., the length of camera takes; *shot scale*, i.e., the distance from the camera to the subject; *shot transitions* (cut, fades, wipes); *camera movement* (pan, tilt, zooms); *camera angle* (high-angle, low-angle, bird's eye, etc.).

These formal features were thought of in the past as not having enough variety to become distinctive of a director. However, recent cinema studies revealed the presence of systematic pat-

terns of one of these formal features (i.e., the shot scale) in the production of Michelangelo Antonioni.[2-3]

Our hypothesis is that the statistical distribution and the temporal pattern of a basic set of frequent formal features, namely *shot duration* and *shot scale*, might act as a fingerprint of a specific director, which enables to track the author's production in time, despite individual style variability. The research questions we tackle are therefore the following:

Q1. Are shot duration and shot scale a sufficient feature set to recognize a director?
Q2. Which is the most distinctive feature for each director?
Q3. Is it possible to determine the movie author using automatically computed shot features?

VISUAL FEATURES AND MOVIE AESTHETICS

As compared to other art forms, research of quantitative analysis of recurrent formal features in film studies has not developed much, probably due to the medium's increased sensory complexity. Back in the 1970s Salt started to manually annotate the presence of formal features,[1] but it is only with the advent of automatic methods that researchers started to systematically investigate the visual and audio dimensions designed by film directors to emotionally impact the audience.[4]

With respect to the visual sphere, different studies focused on conventional MPEG7 color features which include structure, layout, dominant and scalable color. Researchers discovered that saturation and color difference are crucial for mood elicitation in subjects, and as well as color energy and lighting keys, which describe two aesthetic techniques often employed in films: *chiaroscuro* (i.e., high contrast between light and dark areas), and *flat lighting* (i.e., low contrast). Also, light spectral composition used during shooting, namely illuminant, is also important for emotional connotation. Regarding motion dynamics, several algorithms unraveled camera motion parameters to assess pan, tilt, zoom, roll, and horizontal/vertical tracking. This is helpful to distinguish cinematographic shots into different classes such as Aerial, Bird-eye, Crane, Dolly, Establishing, Pan, Tilt, and Zoom. Other motion features such as optical flow, motion maps, and MPEG-7 motion activity, keep into account both global motion and its spatial distribution, without distinguishing between camera and object motion. Motion computed in terms of structural tensor, is also used to evaluate the *visual disturbance* of scenes, with higher values for action films and less expected for dramatic/romantic ones.[5]

Other variables under director's control are depth of field, shot length, and shot scale. Depth of field is the range of acceptable sharpness when a lens focuses on a subject at a distance, and estimating it from single images is still a challenging problem. Shot duration and shot scale are also important formal aspects. Shot boundary detection has been a fundamental step of content analysis since early years of computer vision, and it is well known how it influences audience perception: longer durations relax connotation, whereas shorter shots contribute to a faster pace.[6] Varying camera distance (i.e., shot scale) is a common directing rule used to adjust the relative emphasis between the filmed subject and the surrounding scene.[7]

Previously mentioned features are good indicators of different corpora of films such as drama, comedy, noir, etc.[5] However, to the best of our knowledge, no previous research tackled the problem of automatic attribution of movie authorship. The only attempt in visual arts to recognize authorship has been done on paintings by training a neural network for attributing previously unseen painting to the correct author.[8] Automatic recognition of visual features in fine arts appears also in another work by Gatys et al.,[9] who propose a method to separate (and recombine) content and style, which might be an authorship indicator.

MOVIES AND GROUND-TRUTH DATA

Our dataset includes the complete filmography by 8 different directors whose styles are consensually considered highly unique and distinguishable in film historiography of author cinema: Michelangelo Antonioni, Ingmar Bergman, Robert Bresson, Federico Fellini, Jean-Luc Godard,

Martin Scorsese, Quentin Tarantino, and Bèla Tarr, for a total number of 143 movies directed over 70 years (from 1943 to 2013). The challenges posed by such artistic video content lie in the variety of experimental aesthetic situations, the richness of the scene composition, and the presence of unconventional and symbolic content. The number of representative authors is limited, but good enough to demonstrate the feasibility of authorship recognition. The full list of movies, approximately half black-and-white (b&w), half color, is given on the project website.[10]

Ground truth values for shot scale (GT-scale) have been manually annotated, second by second, by a team of three cinema experts on the full filmographies of 6 directors (Antonioni, Bergman, Fellini, Godard, Scorsese, and Tarr) for a total of 120 movies. On a subset of 77 movies by the same 6 directors (movies indicated in the last column with their Cinemetrics ID in the list on the project website),[10] the GT on shot duration (GT-duration) is also available by *Cinemetrics*,[11] an online application populated by film scholars. For films by Bresson and Tarantino, which account for additional 23 movies, no GT shot features are considered, as they serve as pure testing movies.

FEATURE SETS

Shot duration features

A shot is a series of consecutive frames from a camera take showing a continuous action in time and space. An accurate use of shot duration, and its coupling with motion, allow the filmmaker to control the viewers' eye movements and modulate their attention.[6,12]

We categorize shots into 7 classes according to their duration t_d (s): Very Short (VS), $t_d \in [0, 2)$; Short (S), $t_d \in [2, 4.5)$; Short Medium (SM), $t_d \in [4.5, 7)$; Medium (M), $t_d \in [7, 10)$; Medium Long (ML), $t_d \in [10, 22.5)$; Long (L), $t_d \in [22.5, 40)$; and Very Long (VL), $t_d \in [40, \infty)$. The subdivision is obtained by fitting a log-normal distribution on GT-duration data, thus achieving a finer granularity in proximity of the mode, also considering that the average shot duration for a sample of movies from 1960 to 1985 is 7.0 s, while in contemporary movies decreases to 4.3 s.[6]

Shot duration is a 56-dimensional feature vector obtained by concatenating:

- the shot Duration Distribution (*DDistr*) over the seven classes (7-dimensions);
- the shot Duration Transition matrix (*DTrans*), which describes the probability transition in time of going from any class to any other class (49-dimensions).

Shot scale features

Shot scale i.e., the relative distance of the camera from the main filmed object, affects the emotional involvement of the audience and the identification process of viewers with the characters.[7]

In cinema studies it is usually mapped into 7 categories: *Extreme Close-up* (ECU), *Close-up* (CU), *Medium Close-up* (MCU), *Medium shot* (MS), *Medium Long shot* (MLS), *Long shot* (LS), *Extreme Long shot* (ELS). In practical cases, these are further reduced to 3 families: *Close shots* (CS), *Medium shots* (MS), and *Long shots* (LS), as in Figure 1. Close shots focus on a small area of the scene, such as a face, in such a detail that it almost fills the screen, revealing important details of the plot or character's feelings. In Medium shots actors and the setting share roughly equal space. Long shots show a broad view of the surroundings around the subject.

Shot scale is a 12-dimensional feature vector obtained by concatenating:

- the shot Scale Distribution (*SDistr*) over three classes CS, MS, and LS (3-dimensions);
- the shot Scale Transition matrix (*STrans*), which describes the probability transition in time of going from any class to another class (9-dimensions).

Shot scale feature vectors are built on a second-by-second basis since an individual camera take may contain several scales whenever the camera or the objects in the image are moving. Special cases when more shot scales are found in the same frame are not considered in this analysis.

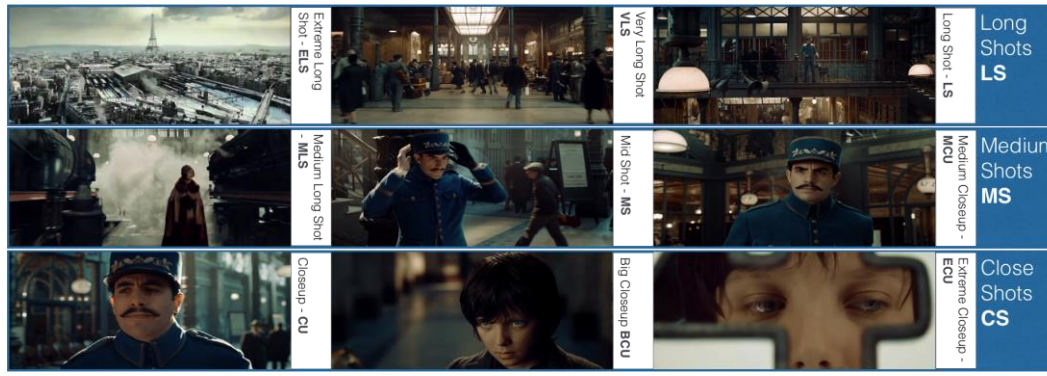

Figure 1. Shot scale classes ranging from Extreme Long Shot (ELS) to Extreme Close-up (ECU), and the three main classes (LS, MS, CS) used in this work (in blue). From *Hugo (Scorsese, 2011)*.

DATA VISUALIZATION

To preliminarily assess the feasibility of automatic authorship recognition, we display and analyze the GT feature data.

Visualizing shot features

The distribution of shot duration (*DDistr*) is shown in Figure 2(a-top) for each of the 6 directors for whom we have GT-duration data (77 movies). Well known Tarr's preference for VL shots is evident. Antonioni and Godard share a tendency to use more ML, L, and VL with respect to Fellini and Bergman, who present similar distributions. Scorsese reveals little use of LS, especially in his recent production.

The distribution of shot scale (*SDistr*) for the same 6 directors (on 120 GT-scale annotated movies) is shown in Figure 2(b-top). Despite small individual differences, such as Antonioni's preference for MS or Scorsese's low standard deviation on scale usage, no director's specific patterns are clearly visible, except the dominance of Close shots on other scales.

Figure 2(a-bottom) represents the transitions between shot duration classes (*DTrans*) for each of the same 6 directors, on their 77 movies annotated with GT-duration, by means of chord diagrams. Classes are arranged radially and the transition probabilities between different classes are represented by connecting arcs. From an observation, it is evident that the probability to move to a shot with different duration is quite high, but often on contiguous classes, meaning that the scene slightly speeds up or decelerates. Since diagrams look very different from author to author, this might be a highly distinctive author feature.

Figure 2(b-bottom) represents the information related to transitions between shot scales (*STrans*) on the 120 movies annotated with GT-scale for the same 6 directors by means of chord diagrams. Since scale is annotated each second, transition probabilities to different scales are pretty small.

Visualizing authorship

To confirm our hypothesis about authorship recognition, we first visualize feature data for each of the 6 directors on the subset of 77 movies with both GT data (GT-scale and GT-duration). By means of t-SNE, a dimensionality reduction technique suited for visualisation of large datasets, we build two and three-dimensional maps in which distances between points reflect similarities. In a reduced 2D feature space (Figure 2(c)) we observe that movies from different authors cluster

together quite well. The ability of the proposed features in separating the 6 authors is better appreciated in the 3-dimensional maps (Figure 2(d) and Figure 2(e)) which represent movies from different directors from convenient view angles.

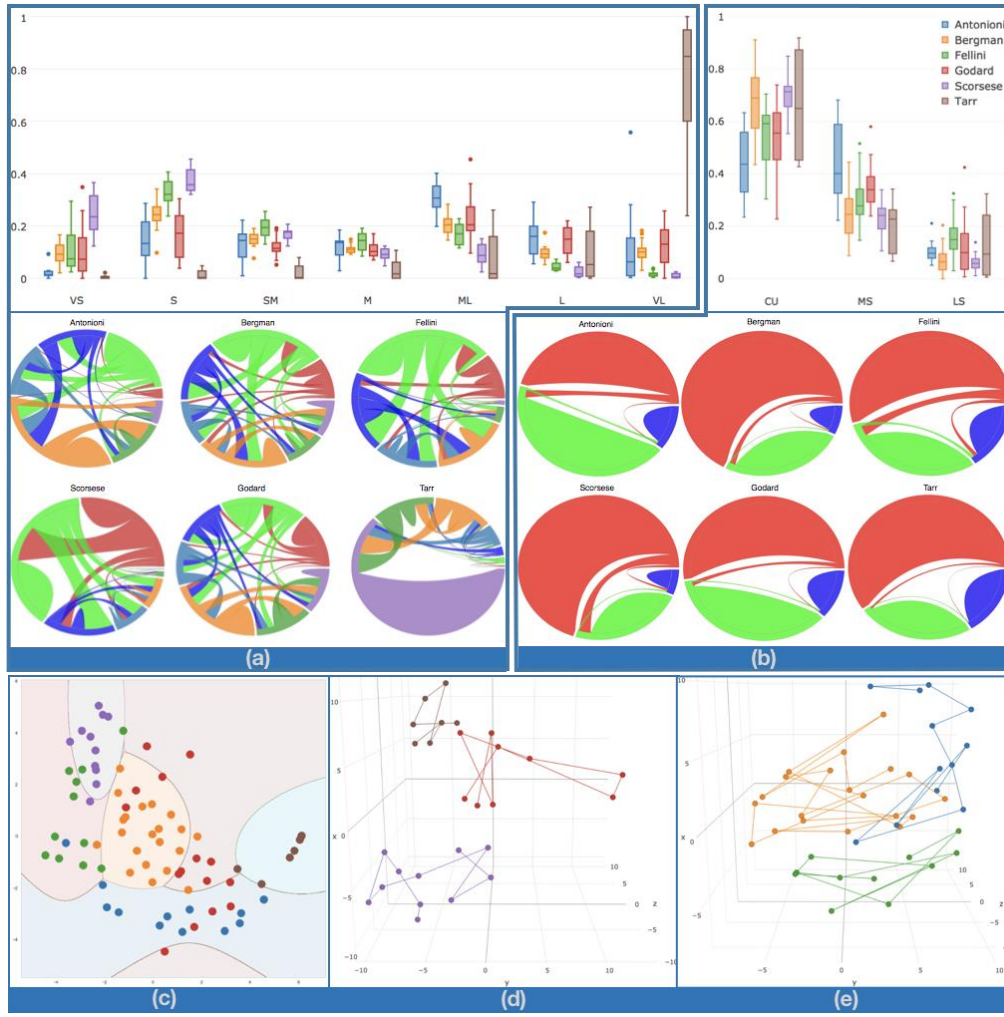

Figure 2. *(a-top)* Shot duration distributions (*DDistr*) for the 6 directors on a total of 77 movies with both annotated GT features (with median, std, variance, and outliers). *(a-bottom) DTrans* on the same movie corpus. In the outer ring the distribution among shot duration classes: VS (red), S (light green), SM (blue), M (light blue), ML (orange), L (dark green), and VL (violet). *(b-top)* Shot scale distributions (*SDistr*) for the 6 directors on 120 movies annotated with GT-scale (with median, std, variance, and outliers). *(b-bottom) STrans on the same movie corpus.* In the outer ring the distribution among scale classes: CS (red), MS (green), and LS (blue). Transitions between classes are represented by the connecting arcs, where the chord widths at the outer ring represent transition probabilities. *(c)* 2D map built with t-SNE on 77 movies annotated with both GT features by Godard (red), Scorsese (violet), Tarr (brown), Antonioni (blue), Bergman (orange), and Fellini (green). 3D maps showing movie clusters on 3 authors, with same colors as before: *(d)* Godard, Scorsese, and Tarr; *(e)* Antonioni, Bergman, and Fellini.

AUTHOR RECOGNITION: FEASIBLE OR NOT?

Author recognition from GT feature data

To answer Q1 (*are shot duration and shot scale a sufficient feature set for recognizing a director?*) we evaluate the performance of different classifiers in attributing movies to the 6 authors on the 77 movies with both GT annotated features, by considering all shot features *DDistr*, *DTrans*, *SDistr*, and *STrans*.

Employing a classical leave-one-out cross-validation i.e., for every author all movies minus one are used for training and the leftovers for testing, we extensively test these classifiers: Nearest Neighbors (KNN), Linear SVM, RBF-kernel SVM, Random Forest, Multi-layer Perceptron (100 hidden nodes and 3 layers), AdaBoost, and Naïve Bayes (NB), under several coarse parameterizations. Best three, which are linear SVM, KNN, and NB, are then accurately tuned by means of a grid search procedure with cross-validation. The fact that SVM, and even more KNN and NB, perform better than other classifiers, is probably due to the relative scarcity of training data. Overall best classification results using both GT features are obtained with the Gaussian Naïve Bayes algorithm; they are presented for each of the 6 authors in Table 1 (white columns) and by the confusion matrix in Figure 5(a) (results in brackets).

Apart from Fellini and Antonioni which show slightly lower recall, all authors are recognized with high accuracy and balanced recall and precision. Obtained results, proving the separability properties of the proposed feature set, are an indicator of the feasibility of automatic recognition of movie authorship by means of automatically computed shot features.

Insights on shot features

We carry out further analysis on single features to better understand their relevance in classification, feature correlation, dependency with respect to the year of production, and most distinctive feature for each director (Q2).

To investigate feature classification abilities when taken alone, the obtained accuracy scores reveal a superior ability in classification by duration features with respect to scale related ones (DDistr=0.701, DTrans=0.636, SDistr=0.272, STrans=0.597).

Figure 3(a) depicts the correlation matrix between all feature components (arranged in blocks). To investigate other possible relations between features and authors, correlations have been computed also for *accessory features*, such as the *year* of production, the total number of movie shots (*#Shots*), the total duration (*#Frames*), and the number of scale changes in the movie (*#SChanges*), which we previously investigated as being responsible for triggering audience's emotional involvement.[13] From the inspections of Figure 3(a) no other conspicuous correlations between feature blocks emerge.

We have also tried to exploit these accessory features in classification, obtaining no significant difference with respect to results in Table 2 (+1% in accuracy). By adding the *year* of production to our feature set, we obtain a slight increase in accuracy of about 2% points.

Since movies have changed dramatically over the last century, with several reflections on film style, especially on the increased shot pace, on motion, and luminance,[6] we also perform a temporal analysis on features. If we restrict the analysis to the shot duration of movies with GT data in Figure 3(b), Scorsese is the only one who significantly decreases his mean shot duration over the years. It is also true that he is the only one (with Tarr, who is obviously an outlier with respect to duration) who lately directed a consistent number of movies. Since there is a natural high correlation between the year of production and the author, with a similar classification process we also experimentally exclude that the classifier learns the author from the year.

Finally, to verify if some features are more relevant than others for recognizing an author, we train a classifier for each director with the one-against-all approach. For assessing how much a certain feature is specific of an author, we exploit the *feature importance measure* returned by

the Random Forest classifier. Figure 3(c) confirms that duration features are in their absolute value more important than those related to shot scale. For all directors, transition features have higher impact than the related distributions. Accessory features (*#Frames*, *#Shots*, and *#SChanges*) are, in absolute terms, of lower importance. In Figure 3(d) importance measures are normalized to the mean across authors, thus highlighting individual differences. Scorsese differs from others for the total number of movie shots. Shot scale distribution is the most specific feature for Antonioni instead, as expected from previous studies,[Q3] while Fellini's most distinguishing feature is the transition pattern between shot scales. Bergman and Godard do not present any peculiar feature, and this is probably why their authorships are sometimes confused in classification, as shown in Figure 5(a). As expected, Tarr is characterized by his unique use of long takes which impacts on the duration distribution.

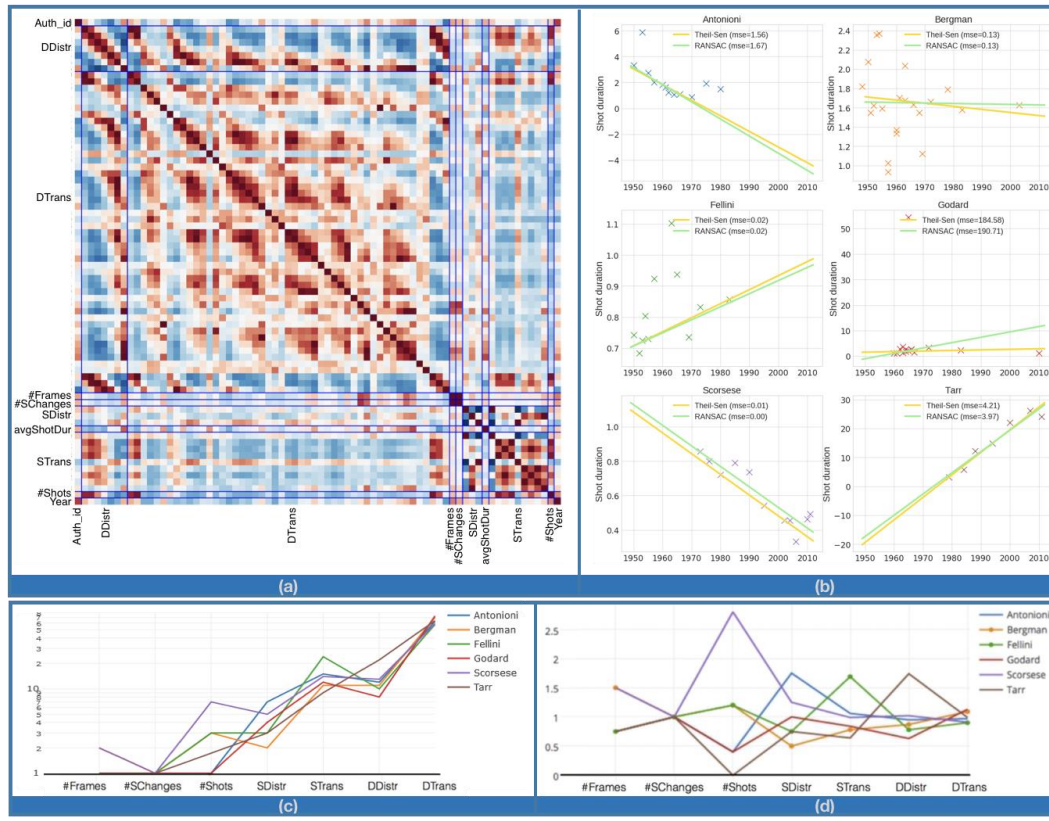

Figure 3. (a) Correlation matrix between all feature components. (b) Trends of shot duration for the 6 directors on the 77 movies with both GT data (obtained by Ransac and Theil-Sen methods). (c) Importance measure of single features for each of the 6 director on the same corpus, in absolute value and (d) normalized to the mean across authors.

AUTOMATIC SHOT FEATURE ANALYSIS

After the analysis with GT features, we perform author recognition by means of fully automatic procedures. Shot transitions are detected by existing methods, while we use our deep learning approach to derive shot scale.[14] Eventually we compare performance on author recognition with results obtained using GT features.

Shot duration features

For segmenting videos into shots we evaluate the algorithm by Apostolidis et al. and the one by Adami et al. which has been recently extended to identify cuts.[15,16] The first method detects abrupt and gradual transitions based on frame similarity computed by means of both local (SURF)

and global (HSV histograms) descriptors, while the second exploits histogram information and decision criteria derived from statistical properties of cuts, dissolves, and wipes. To select the algorithm, we test both methods on *Il Deserto Rosso (Antonioni, 1964)*, which counts more than 700 shots.

Scores are computed with respect to a manually annotated GT at frame level, confirming the efficiency of both methods (Apostolidis et al.: Pr.=0.935, Rec.=0.910, F1=0.922; Adami et al.: Pr.=0.892, Rec.=0.970, F1=0.929). Since the first method is slower than the recent implementation by Adami et al., named *scdetector*, which works in real-time, we use the latter for producing *DDistr* and *DTrans* features on the whole set of 143 movies. We show in Figure 4 the duration distributions obtained by *scdetector* on *Le notti di Cabiria (Fellini, 1957)* and *Gangs of New York (Scorsese, 2002)* (in red), overlaid to GT ones (in blue).

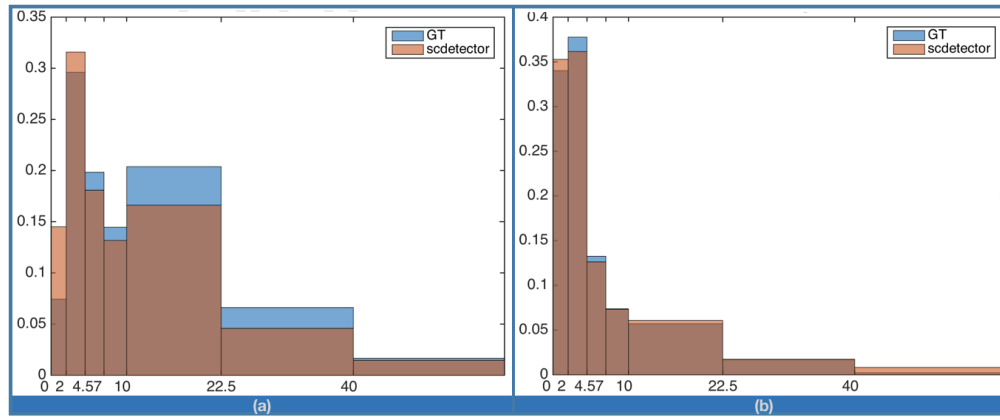

Figure 4. Automatically computed shot duration distributions (in red) and GT ones (in blue) for (a) *Le notti di Cabiria (Fellini, 1957)*, and (b) *Gangs of New York (Scorsese, 2002)*.

For most movies, automatically derived *DDistr* show little deviation from the GT distributions, with few problems only on Godard's production. Once computed shot boundaries, *DTrans* are easily derived with no substantial difference with respect to those in Figure 2(a-bottom).

Shot scale features

We use a novel approach for shot scale estimation based on deep learning techniques.[14] Exploiting the GT-scale dataset using 55 movies for training and the other 65 for testing, we compare three Convolutional Neural Networks with increasing capacity (AlexNet, GoogLeNet, and VGG16) trained by an extensive hyperparameter selection process.[17-18] To counteract possible fluctuations of the output, we exploit the temporal correlation of the shot scale by means of a smoothing moving median window of 3s.

Among tested networks, VGG-16 performs best with an overall precision of 94%, recall 94%, and accuracy 94% by fine-tuning the whole network. For most movies, automatically derived scale distributions by VGG-16 show little deviation from the corresponding GT distributions. Once computed the shot scale for each movie, *STrans* features are easily derived with no substantial difference with respect to those in Figure 2(b-bottom).

FULLY AUTOMATIC AUTHOR RECOGNITION

To answer Q3 (*is it possible to robustly determine the authorship of a movie using automatically computed shot features?*) we repeat the experiments on author attribution with automatically extracted features.

Performance on author attribution are first evaluated on the same set of 120 movies by the 6 directors (55 movies for testing, 65 for training - see the list on the project website -) in order to

compare with attribution computed by means of GT features. By employing a leave-one-out cross-validation, we compute precision, recall, and accuracy, exploiting the Gaussian Naïve Bayes classifier. Results are presented for each author in Table 1 (gray columns) and by the confusion matrix in Figure 5(a) (in brackets results obtained by using GT feature values). Quite surprisingly there is no drop of performance in fully automatic recognition with respect to the classification achieved by GT shot features, even though tests are conducted on a different number of movies (65 instead of 77). This implies that automatically computed features are in general robustly extracted. Using distributions and transition matrices as features has the important advantage that false positive and false negative for each class tend to compensate, de-facto softening the impact of errors on overall classification accuracy. Therefore, even in case of slight differences with respect to GT, still shot feature distributions and transition patterns change from author to author in a way that is strongly related to the individual author style.

For the second test we add the filmographies by two authors, Tarantino and Bresson, which account for additional 23 full movies, never seen before by the shot scale detector during its training phase. We then repeat the experiments on author attribution with automatically extracted features on the full corpus by 8 different directors (143 movies in total), using a leave-one-out cross validation, i.e., for every author all movies minus one are used for training and the leftovers for testing. Obtained results, shown in Table 2 and in Figure 5(b), are again clearly far above the chance level (12.5%).

Table 1. Comparison on authorship recognition by GT vs. automatic features on 6 directors.

	PRECISION		RECALL		F1-SCORE		MOVIES	
AUTHOR	**GT-feat**	**Aut-feat**	**GT-feat**	**Aut-feat**	**GT-feat**	**Aut-feat**	**GT-feat**	**Aut-feat**
Antonioni	0.75	0.60	0.50	0.67	0.60	0.63	12	9
Bergman	0.80	0.85	0.76	0.81	0.78	0.83	21	21
Fellini	0.70	0.71	0.64	0.71	0.67	0.71	11	7
Godard	0.57	0.50	0.87	0.50	0.68	0.50	15	12
Scorsese	0.80	0.77	0.73	0.91	0.76	0.83	11	11
Tarr	0.83	1.00	0.71	0.60	0.77	0.75	7	5
Avg/Tot	***0.74***	***0.73***	***0.71***	***0.72***	***0.71***	***0.72***	***77***	***65***

Table 2. Automatic authorship recognition by automatically computed features on 8 directors (Naïve Bayes, leave-one-out cross-validation).

	PRECISION	RECALL	F1-SCORE	MOVIES
AUTHOR	**Aut-feat**	**Aut-feat**	**Aut-feat**	**Aut-feat**
Antonioni	0.60	0.67	0.63	9
Bergman	0.89	0.81	0.85	21
Fellini	0.56	0.71	0.63	7
Godard	0.50	0.58	0.54	12
Scorsese	0.75	0.82	0.78	11
Tarr	1.00	0.60	0.75	5
Tarantino	1.00	0.89	0.94	10
Bresson	1.00	0.92	0.96	13
Avg/Total	***0.80***	***0.77***	***0.78***	***88***

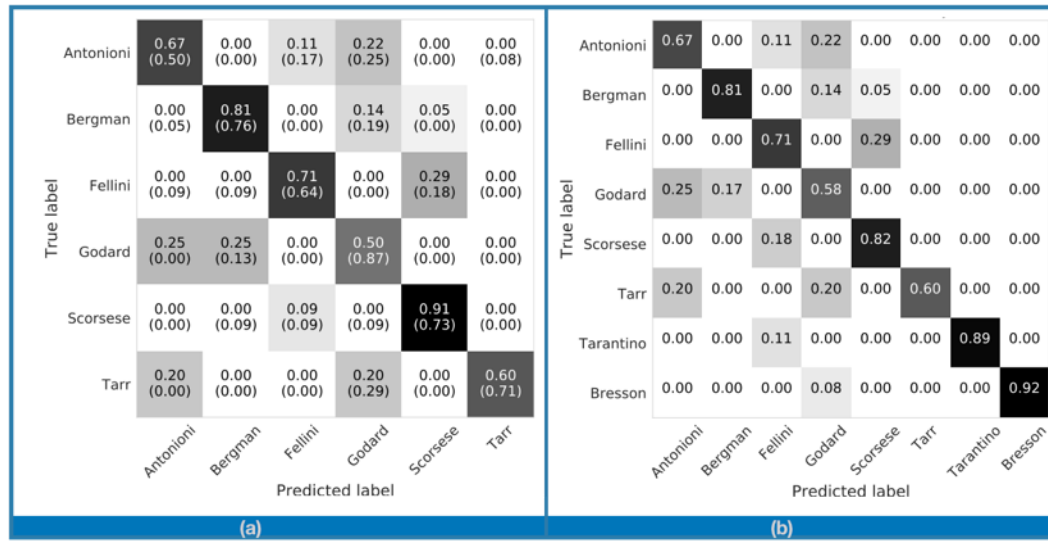

Figure 5. Confusion matrices: (a) Authorship attribution by automatic features on 6 directors compared with attribution computed by GT features (in brackets); (b) Authorship attribution computed with fully automatic features on 8 directors.

DISCUSSION

We think about works of art as consciously designed and carefully executed to elicit meanings and emotions. For example, it is true that shot scales are meticulously planned one by one in a film. But a shot scale distribution of a whole film is something that would not occur to any filmmaker. Here is what Antonioni says:

> *"When I am shooting a film I never think of how I want to shoot something; I simply shoot it. My technique [...] is wholly instinctive and never based on a priori considerations."*[20]

Yet, this study shows that statistical patterns of formal features are systematically capable of distinguishing between individual authorial styles. Even though these patterns cannot be conceived as consciously designed, they underlie intentionally crafted features. Whether this relationship is correlational (e.g., narrative features involve specific visual techniques) or perceptual and aesthetic (e.g., a certain aesthetic effect is obtained at a certain frequency of a technical feature value, like eliciting empathy with characters is obtained by frequently shifting from any shot scale to close-up) needs further investigation: stylistic, psychological and statistical.

It is important to notice that sequence patterns are at least as characteristic of individual styles as their distributions. We think that the reason is that sequence patterns characterize the dynamics of film form, which has a closer relationship to narration than overall distributions. The sequences of shot duration and shot scale might be closely related with such fundamental aesthetic patterns as rhythm, regulation of arousal and of emotional involvement. Since the dynamics of these aesthetic effects are the key to regulate dramatic tension in films, which in turn is the key to keep the viewer's attention, it seems probable that sequence patterns are more under conscious authorial control than overall values of the same formal features. However, individual styles show a big variety in terms of which feature is the most characteristic of them as compared to the average values across all authors. And sometimes no single feature is characteristic for an author's work as with Bergman and Godard. However, this does not affect the recall rate of their films, which suggests that taking these dimensions together is very robust in the attribution task.

Despite the large database of 143 movies and the consequent huge number of automatically analyzed frames (in the order of 1 million) we are aware that a limitation is the small quantity of the considered directors. The reduced number of artists might have a strong influence on the perfor-

mance, which suggests that a further improvement should be directed towards expanding the director set. However, this work experimentally confirms the hypothesis that the use of a few formal features is sufficiently distinctive, and remains constant enough in one author's production, to enable an automatic attribution of the authorship. The fact that by using only shot duration and shot scale it is possible to distinguish among 8 different authors suggests that such an analysis can easily scale up to more directors and larger databases. We are confident that adding other formal features, both perceptual and technical (e.g., depth of field, camera motion, angle, etc.), would increase the accuracy of identification.

Feature learning is nowadays a valid alternative to the use of hand-crafted features. However, in this case learning features directly from data would need an enormous number of examples to isolate the distinctive features. For example, to discover artists' characteristic visual features, the CNN by van Noord et al. works on the Rijksmuseum Challenge database:[8] more than 110,000 digitalized artworks by more than 6,000 painters. Conversely, working with formal features has the non-trivial advantage of allowing direct testing of hypothesis coming from cinema studies and psychology, thus broadening the scope and the interest of this work to more disciplines.

CONCLUSION

This work represent an example of cross-disciplinary research inspired from studies in cinema about the peculiar use of formal features by different directors.[2] Future work will aim at better investigating the effect of such features on viewers from the neuroscience perspective,[20] and linking it with empathy-related processes felt during the cinematic experience.[7] Last, since using a subdivision of shot scale in three classes could limit the ability of author attribution, another possible extension should aim at obtaining at distinguishing more intermediate scales.

ACKNOWLEGMENTS

Authors M. Svanera, M. Savardi, and S. Benini equally contributed to this research work.

REFERENCES

1. B. Salt, *Moving Into Pictures: More on Film History, Style, and Analysis*. Starword, 2006. [Online]. Available: http://books.google.it/ books?id=wW8cAgAACAAJ
2. A. B. Kovács, "Shot scale distribution: an authorial fingerprint or a cognitive pattern?" *Projections*, vol. 8, no. 2, 2014.
3. S. Benini, M. Svanera, N. Adami, R. Leonardi, and A. B. Kovács, "Shot scale distribution in art films," *Multimedia Tools and Applications*, vol. 75, issue 23, Dec 2016.
4. H. L. Wang and L. F. Cheong, "Affective understanding in film," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 6, pp. 689–704, June 2006.
5. Z. Rasheed,Y. Sheikh, and M. Shah,"On the use of computable features for film classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 52–64, Jan 2005.
6. J. E. Cutting, K. L. Brunick, J. E. DeLong, C. Iricinschi, and A. Candan, "Quicker, faster, darker: Changes in hollywood film over 75 years," *i-Perception*, vol. 2, no. 6, pp. 569–576, 2011. [Online]. Available: http://dx.doi.org/10.1068/i0441aap
7. K. Bálint, T. Klausch, and T. Pólya, "Watching closely*: Spatial distance influences theory of mind responding in film viewers*," *Journal of Media Psychology*: theories, methods and applications. ISSN 1864-1105 vol. 0, no. 0, pp. 1–10, 2016. [Online]. Available: http://dx.doi.org/10.1027/1864- 1105/a000189
8. N. van Noord, E. Hendriks, and E. Postma, "Toward discovery of the artist's style: Learning to recognize artists by their artworks," *IEEE Signal Processing Magazine*, vol. 32, no. 4, pp. 46–54, July 2015.

9. L. A. Gatys, A. S. Ecker, and M. Bethge, “A neural algorithm of artistic style,” *arXiv*, Aug 2015. [Online]. Available: http: //arxiv.org/abs/1508.06576
10. “Who is the director of this movie?” 2017, [Online; accessed 18-May-2018]. Available: http://projects.i-ctm.eu/en/project/who-director-movie-shot-analysis-art-films
11. “Cinemetrics: Movie measurement and study tool database,” 2017, [Online; accessed 18-May-’18]. Available: www.cinemetrics.lv
12. T. J. Smith, “Attentional theory of cinematic continuity,” *Projections: The Journal for Movies and Mind*, vol. 6, no. 1, pp. 1–27, 2012.
13. L. Canini, S. Benini, and R. Leonardi, “Affective analysis on patterns of shot types in movies,” in *Image and Signal Processing and Analysis (ISPA), 2011 7th Int. Symposium on*. IEEE, 2011, pp. 253–258.
14. M. Savardi, A. Signoroni, P. Migliorati and S. Benini, "Shot Scale Analysis in Movies by Convolutional Neural Networks," *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, 2018, pp. 2620-2624. doi: 10.1109/ICIP.2018.8451474
15. E. Apostolidis and V. Mezaris, “Fast shot segmentation combining global and local visual descriptors,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 6583–6587.
16. N. Adami and R. Leonardi, “Identification of editing effect in image sequences by statistical modeling,” in *in Proc. of the 1999 Picture Coding Symposium (PCS 1999)*, 1999, pp. 157–160.
17. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012, pp. 1097–1105.
18. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed ,D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1–9.
19. C. T. Samuels, “Interview with Michelangelo Antonioni in Rome,” July 29th 1969, [Online; accessed 11-Sep-2018]. Available: http://www.michelangeloantonioni.info/2013/05/04/interview-with-michelangelo-antonioni-by-charles-thomas-samuels/
20. G. Raz, M. Svanera; N. Singer, G. Gilam, M. Bleich Cohen, T. Lin, R. Admon, T. Gonen, A. Thaler, R. Y. Granot, R. Goebel, S. Benini, G. Valente, "Robust inter-subject audiovisual decoding in functional magnetic resonance imaging using high-dimensional regression", in Neuroimage, vol. 163, pp. 244-263, December 2017.

ABOUT THE AUTHORS

Michele Svanera obtained his B.Sc, M.Sc, and PhD in Telecommunication Eng. at University of Brescia. During his PhD he moved across AI and Cognitive Neuroscience to understand brain activities under movie stimulus, visiting important brain centres in Maastricht and Tel Aviv. Since May 2017 he joined Lars Muckli’s Lab, University of Glasgow, as a post- doc, working on high-field fMRI and localisation of non-feedforward sources in primary visual cortex.

Mattia Savardi received his M.Sc. degree in Communication Technologies and Multimedia (cum laude) at the University of Brescia in 2016. He is currently a Ph.D student at the University of Brescia since November 2016. His efforts are on semantic and multimedia content analysis and biomedical signal and imaging applications using Computer Vision and Deep Learning techniques.

András Bálint Kovács is professor and founding chair of the Dept. of Film Studies at ELTE University, Budapest. He teaches history, analysis and cognitive theory of cinema. His research includes psychological studies of viewer’s reactions to causal connections in filmic narratives, psychology of perception of shot scales, and statistical analysis of film

style. He has been visiting professor at Univ. of Stockholm, Université dela Nouvelle Sorbonne, École Normale Supérieure, and UC San Diego.

Alberto Signoroni received M.Sc. in Electronic Engineering 97 and Ph.D. in Information Engineering 01 from the University of Brescia, Italy. Currently he is an Assistant Professor with the Information Engineering Department at the University of Brescia. His research interests include 3D computer vision and geometry processing, biomedical image analysis, multidimensional and hyperspectral image processing.

Sergio Benini received both his MSc in Electronic Engineering in 2000 (cum laude) and PhD in Information Engineering (2006) from the University of Brescia, IT. Between ’01 and ’03 he was with Siemens Mobile Communications. During his Ph.D. he spent one year in British Telecom UK. Since 2005 he is Assistant Professor at the University of Brescia. In 2012 he co-founded Yonder http://yonderlabs.com, a spin-off in NLP, ML, and Cognitive Computing.