# PLOTSHOT: Generating Discourse-constrained Stories around Photos

**Rogelio E. Cardona-Rivera**[1,2] and **Boyang Li**[2]
[1]Liquid Narrative Group, Department of Computer Science, North Carolina State University
[2]Disney Research
recardon@ncsu.edu, albert.li@disneyresearch.com

**Abstract**

Story generators typically adopt a pipelined model of generation wherein fabula structure is decided independently and prior to discourse structure. In this paper, we propose a novel story generator, PLOTSHOT, capable of reasoning over discourse materials during fabula generation such that these materials meaningfully constrain the development of a causally and intentionally coherent story. PLOTSHOT incorporates user-supplied photographs as optional story states through an oversubscription planning paradigm. Further, to leverage existing work on planning-based models of generation, we present a technique to compile the photo story planning problem to classical narrative planning. Our system attempts to maximize quality of an illustrated story by analyzing the affinity between a photo and the action it is meant to depict. An evaluation of generated artifacts shows advantage over heuristic baseline techniques.

## Introduction

Research in story generation has advanced a bipartite representation of narratives (Young 2007). One part is termed the *fabula*, which is the conceptualization of the story world including actants that exist and events that transpire. The other part is termed the *discourse*, which includes elements responsible for the story's telling. Most existing story generation work (for reviews see Gervás 2009; Ronfard and Szilas 2014) has focused on a fabula-driven pipeline. This approach commits to a fabula by simulating the story world as a sequence of events, and then creates a discourse by selectively presenting those events; fabula is generated independently of and prior to the discourse.

However, to create captivating fabula, it is important to consider discursive constraints. For instance, a movie production may choose to include a car chase scene because of its visual appeal. As a result, the fabula must adapt and provide causal and motivational justification for the car chase in order to maintain story coherence. This process can be understood as a discourse placing constraints on a fabula. Further, there may be multiple cinematic scenes that have different visual appeal and require different story structures. A story generator should select visual scenes to optimize for and balance visual appeal with story coherence.



We propose one such story generator, which we call PLOTSHOT. This system considers optional discourse goals, which represent available discourse materials, and attempts to incorporate them into the generated story. Our formulation combines classical narrative planning with oversubscription planning (Smith 2004). The former affords reasoning about authorial intent, i.e. achieving specific story outcomes and intermediate states. The latter affords reasoning about optional intermediate states and a hard cost budget to fit stories to fixed-length formats, e.g. feature-length movies or booklets. To solve the hybrid problem, we present a technique to select amongst available discourse materials and compile the novel problem into a classical problem in order to leverage existing work in planning-based narrative generation.

The specific domain of application for this paper is generating illustrated stories around a set of user-supplied photos. We represent photos via scene graphs, semantic networks that describe the photos' content. As a photo depicts the current world state, it can be displayed only when all of its content has been established. Since there are often multiple locations where a photo may be displayed, we present a photo placement strategy based on the affinity between photos and story actions. Our evaluation of generated photo booklets demonstrates the effectiveness of the affinity-based strategy over heuristic baseline techniques.

**Contributions** In this paper(1) we propose a novel challenge for selective incorporation of discourse materials into fabula generation, which is beyond the capabilities of state-of-the-art narrative planning systems; (2) we formulate the challenge as a hybrid classical/oversubscription planning problem and present a technique to compile the hybrid problem into a classical problem; (3) we define a metric to evaluate the affinity between a photo and the story event it depicts, and (4) demonstrate that an affinity-based photo placement strategy is significantly better than baselines.

## Related Work

Fabula-driven approaches have been the dominant model in automated story generation (e.g. Aylett, Dias, and Paiva 2006; Li 2015; Teutenberg and Porteous 2015; Barot, Potts, and Young 2015). However, there are a few exceptions where discourse reasoning substantially impacts fabula.

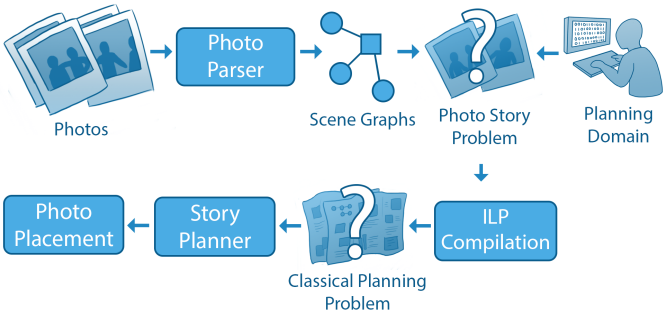


Figure 1: The PLOTSHOT system architecture. Boxes with text represent algorithmic modules.

The Slant story generation system (Montfort et al. 2013) detects aspects of the verbs during discourse generation and determines the fabula's match to different genres. After a specific genre has been identified, additional constraints are posed to the fabula generator. Unlike their work, our system works with discourse constraints supplied by users in a novel planning formulation. Piacenza et al. (2011)'s video recombination system is very similar to our work. They learn some basic semantic units of video and a Markov transition model between the units. All actions in the fabula plan must have a corresponding video clip. The planner sends requests for video clip. If a request is denied, the story can be replanned. In comparison, we allow actions without photos and describe such actions with text. In our formulation, discourse materials are optional goals that the planner explicitly attempts to achieve, not requested after a plan is created. Radiano et al. (in revision) select and fit photos to a given story template instead of dynamically adapting stories to suit photos.

Some interactive story systems aim to satisfy mandatory intermediate story goals (Riedl 2009); these are similar to ours as they all need to work with events that must happen or must have happened, given discourse events that have been observed. Robertson and Young (2014) propose an algorithm for modifying early events that are unknown to the player but are causally related to later, known events. These early events can be replaced as long as preconditions of the known events can be satisfied. Tomai (2014) developed a system that is free to assert events that may reasonably have happened at the same time or slightly before observed player actions. In addition to mandatory goals, our work also supports optional intermediate goals that may be abandoned if they are deemed too costly.

## The PLOTSHOT System

PLOTSHOT is a narrative generation system for photo-illustrated stories. Its architecture is illustrated in Figure 1. PLOTSHOT relies on a Photo Parser module to parse a photo into a set of logic literals, which form an optional discourse goal. A photo story planning problem contains optional discourse goals and a planning domain designed by domain engineers. The ILP Compilation module selects optional discourse goals to create an approximate classical problem that can be solved by an off-the-shelf narrative planner. After narrative planning, a Photo Placement module places photos in the story. The process employs a measure of compatibility between story actions and photos, defined as *photo-action affinity*. We describe these modules in following sections.

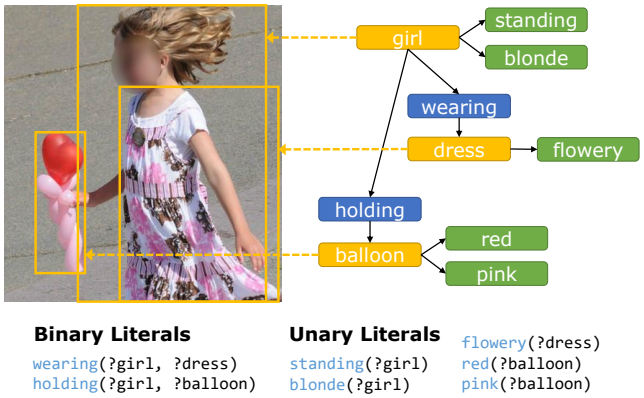


Figure 2: An example of a scene graph and its grounding parsed from an image (top) and the extracted logical literals (bottom). The scene graph contains objects (in yellow), relations (in blue) and attributes (in green).

### Representing Photos

Computationally parsing the semantic meaning of images is an open and difficult challenge. Recently, Krishna et al. (2016) proposed parsing images as *scene graphs*. A scene graph is a directed graph consisting of *objects* (e.g., people, places, things), *relationships* between objects (e.g., on-top-of, wearing), and *attributes* of objects (e.g., color, shape). Each object is grounded through a bounding box.

We assume that we have accurately parsed scene graphs from user-supplied photos, and that these graphs preserve identities for actors across the photos using techniques like that of Dai et al. (2015). Such a parser is beyond the scope of this work. The main reason for adopting this representation is its active use in computer vision research.

To bridge with the representation used by narrative planners, we straightforwardly convert a scene graph to a set of logic literals: relations can be converted to binary literals and attributes to unary literals. We opt to treat objects in the extracted literals as variables, so that they can be bound with planning problem objects. Figure 2 shows an example of an example scene graph and logic literals extracted from it.

There is a potential for domain disparity between the scene graphs and our planning domain. That is, some literals in the photos (e.g., hair color as in Figure 2) may not be specified anywhere in the planning domain definition. As a result, they can never be achieved. To get around this issue, we remove from the parsed literals the predicates not defined in the planning domain. A photo with no literals left cannot be used in the story.

### Photos as Optional Goals in Planning

In this section, we formulate the photo-story problem based on the needs for story generation from photos.

**Background** Young et al. (2013) model a story as a plan, or a sequence of actions that transforms an initial state $s_0$ to achieve a set of goal literals $g_\infty$. Each action $a$ is associated with a set of preconditions $\text{PRE}(a)$, effects $\text{EFF}(a)$, and a cost $c(a)$. Effects are bipartite: positive literals are recorded in an add list, and negative literals are recorded in a delete list. An action is applicable in a state when its preconditions are true in that state. After an action is executed, literals in the delete list are removed from the state and literals in the add list are added to it. In classical planning, goal satisfaction is a hard constraint: a plan is valid only if all goal literals are made true. Conversely, cost is a soft constraint being optimized.

Over-subscription planning (OSP) (Smith 2004) is a different formulation. An OSP problem contains a set of optional goals and a maximum cost $C_{max}$. A valid solution must cost less than $C_{max}$. An example OSP problem is for a robot to collect rewards from different locations subject to a finite amount of fuel. Different locations can have different benefits and different routes consume different amounts of fuel. We want to maximize the sum of benefits before exhausting the fuel. Unlike classical planning, in OSP goal satisfaction is a soft constraint and cost is a hard constraint.

**The Use Case** We formulate the *photo story planning problem* as a hybrid of OSP and classical planning with both optional and mandatory goals. This is driven by our use case of story generation around user-supplied photos.

Photos in our stories are naturally optional because users usually have many more photos than we can put into a story, so some selection of photos is necessary. A photo is considered to depict a state of the underlying story world. To maintain story coherence, all facts that are present in a photo must be causally established (but not always explicitly presented) before the photo can be shown.

Mandatory goals arise from the need to specify story outcomes and the need to represent *authorial intentions* (Riedl 2009), which indicate intermediate story states through which all valid stories must pass. For example, in the story world of Cinderella, an intermediate story state could be that Cinderella leaves her shoe at the palace at some point. One condition for the story outcome is she gets her shoe back. The ability to specify authorial intentions and outcomes grants effective control over the narrative arc to the narrative domain engineer. A upper bound on cost allows us to fit a story into a fixed-length format such as a poster.

**The Photo Story Planning Problem** Formally, a photo story planning problem is a tuple $\langle S, s_0, g_\infty, G^o, G^M, A, \sigma(\cdot), c(\cdot), C_{max}, b(\cdot)\rangle$. We first explain notations related to state trajectories. $S$ is the set of possible states. $s_0 \in S$ is the initial state. $A$ is a set of possible actions. An action $a \in A$ can be applied to a state $s \in S$ if its preconditions $\text{PRE}(a) \in s$. $\sigma(\cdot)$ denotes the deterministic transition function; we let $s' = \sigma(s, a)$ denote that applying $a$ in $s$ results in a new state $s'$. For a given action sequence $p = [a_1, \ldots, a_n]$, we let $\tau(s, p)$ denote the state trajectory resulted from applying $p$ sequentially to $s$. That is, $\tau(s, p) = [\sigma(s, a_1), \sigma(\sigma(s, a_1), a_2), \ldots]$.

The following notations define plan goals. Goals in the problem include the story goal $g_\infty$, a set of optional goals $G^o$, and a set of mandatory intermediate goals $G^M$. Each goal $g \in G^o \cup G^M \cup \{g_\infty\}$ is a set of logic literals. An intermediate goal $g^i$ is achieved by an action sequence $p$ if any state in $\tau(s_0, p)$ contains all literals in $g^i$. The final story goal $g_\infty$ is achieved by $p$ if the last state in $\tau(s_0, p)$ contains all literals in $g_\infty$.

The following notations define plan quality. The function $c(a)$ measures the cost of action $a$. $C_{max}$ is the limit on total cost. A benefit value $b(g^o)$ is defined for each optional intermediate goal $g^o \in G^o$, which represents the aesthetic quality of the photo. In this paper, we assume aesthetic quality of a single photo is given to us as a single number, as much research has been done on its automatic assessment (e.g., Joshi et al. 2011; Morgens and Jhala 2013). In later sections, we consider another aesthetic concern of the connection between photos and actions. A valid solution to the problem is a sequence of actions $p$ that can be applied in the initial state, achieves all of $G^M$ and $g_\infty$, and has a total cost $\sum_{i=1}^{k} c(a_i) \leq C_{max}$. Among valid solutions, we seek one that maximizes the sum of the benefits of all achieved optional goals.

Domshlak and Mirkis (2015) argue that OSP cannot reduce to classical planning without incurring a loss of plan quality. Mixing optional and mandatory goals further adds to the difficulty of the problem. However, compiling our photo story planning problem to classical narrative planning allows us to benefit from decades of extensive research on the latter problem. In the next section we present an approximate compilation technique.

## The Compilation: Choosing Mandatory Goals from Optional Goals

We propose a technique for selecting some optional goals and convert them to mandatory goals in classical planning. This is similar to Smith (2004) but the difference is that we consider both optional and mandatory goals in the selection.

Goal selection can be thought of as traversing a directed graph where each node represents a goal. Each directed edge has a weight, representing the additional cost needed to reach the second goal after achieving the first goal. A valid path starts at the initial state, ends at the story goal, and visits every mandatory intermediate goal and some optional intermediate goal. We seek a valid path that maximizes the total benefit of visited optional goals. This formulation is similar to the traveling salesman problem and its variant, the orienteering problem, but also different due to the existence of both optional and mandatory components. After we find the optimal path, all optional goals on the path are turned to mandatory goals in the planning stage.

We use integer linear programming (ILP) to solve this problem. Although ILP is NP-complete, several fast off-the-shelf solvers exist. For the ILP problem, we introduce vertex variables $V_i \in \{0, 1\}, \forall i \in \{1, \ldots, n\}$ and edge variables $E_{ij} \in \{0, 1\}, \forall i, j \in \{1, \ldots, n\}$. The vertex variables include: the initial state $V_0$, the story goal $V_n$, $k$ optional intermediate goals $V_1, \ldots, V_k$, and $n - 1 - k$ mandatory intermediate goals $V_{k+1}, \ldots, V_{n-1}$. Edge variables $E_{ij}$ and $E_{ji}$ represent two oppositely directed edges between every

vertex pair $V_i$ and $V_j$, with the exception that $V_0$ does not have incoming edges and $V_n$ does not have outgoing edges. Setting an edge variable or a vertex variable to 1 puts the edge or the vertex on the optimal path. The edge weight $c_{ij}$ reflects the cost incurred by going from $V_i$ to $V_j$. If the $i^{\text{th}}$ vertex is optional, its benefits $b_i$ is greater than zero; otherwise $b_i = 0$. The ILP maximizes the following objective function:

$$\max \sum_i b_i V_i$$

while respecting the following constraints:

- $\sum_i \sum_j c_{ij} E_{ij} \leq C_{max}$: the total cost of the plan is no more than a predefined maximum.
- $V_i = 1, \forall i \in \{1, k+1, \ldots, n\}$: the initial state, the goal state, and all authorial goals must be on the path.
- $\sum_j E_{ji} = V_i, \forall i \neq 1$: every selected vertex has an incoming edge, except the initial state.
- $\sum_j E_{ij} = V_i, \forall i \neq n$: every selected vertex has an outgoing edge, except the goal state.

We prevent cycles by introducing auxiliary variables $U_i, \forall i \geq 2$ and the following constraints (Miller, Tucker, and Zemlin 1960):

- $2 \leq U_i \leq n, \forall i \in \{2, \ldots, n\}$
- $U_i - U_j + (n-1)E_{ij} \leq (n-2), \forall U_i, U_j, i \neq j$

Finally, we account for mutual exclusions between goals. It is likely that users provide multiple photos depicting similar underlying content, such as photos of the same subjects in slightly different poses or angles. The literals from the scene graphs of the similar photos will be equivalent. To avoid repetition, we only allow one such photo in our story. This is achieved by the mutex conditions. A mutex set $M$ contains a number of vertices such that only one in $M$ can be selected in the path. That is,

- $\sum_i V_i \leq 1, \forall i \in M$

Unlike Smith (2004) who estimated edge costs based on a manual sensitivity analysis of plan cost, we estimate the edge cost automatically via the use of relaxed plans. A relaxed plan is a directed layered graph that is a subgraph of a planning graph (Blum and Furst 1997). A relaxed plan starts in state $s$ and reaches $s'$ by ignoring the delete lists of all actions; the state contains a monotonically non-decreasing set of literals as more actions are applied to it. The length of a relaxed plan that solves a planning problem represents an estimate of the actual cost to solve said problem. In our case, we use relaxed plan to estimate the distance $c_{ij}$ between the nodes of our graph. We consider three cases:

1. The distance from $V_0$ to an intermediate node $V_i$ is denoted by $c_{0i}$ and is computed with the relaxed plan from the initial state $s_0$ to the $i^{\text{th}}$ intermediate goal.
2. The distance $c_{ij}$ from an intermediate node $V_i$ to another intermediate node $V_j$ is computed as follows. We first compute a relaxed plan $R_{0i}$ from the initial state $s_0$ to the $i^{\text{th}}$ intermediate goal. Following García-Olaya, de la Rosa, and Borrajo (2011), we compute the state at $V_i$ by applying the relaxed plan's actions in sequence, applying both add lists and delete lists and ignoring preconditions. We denote the state obtained from applying $R_{0i}$ as $s^R_{0i}$. We then compute a relaxed plan from $s^R_{0i}$ to the $j^{\text{th}}$ intermediate goal and estimate the distance $c_{ij}$ accordingly.
3. The distance $c_{in}$ from an intermediate node $V_i$ to the goal node $V_n$ is computed with the relaxed plan from $s^R_{0i}$ to the story goal $g_\infty$.

We use the ILP to compile the photo story problem to a classical planning problem. All optional discourse goals that are visited by the optimal path become mandatory intermediate goals. Unvisited discourse goals are discarded. After the compilation, the problem can be solved by any off-the-shelf narrative planner. If the optimal plan found exceeds the cost limit $C_{max}$, the optional discourse goal ordered last in the plan is removed and planning is attempted again. In this paper, we use the Glaive narrative planner (Ware and Young 2014), in which all actions performed by story characters must serve a character intention but not all character intentions succeed.

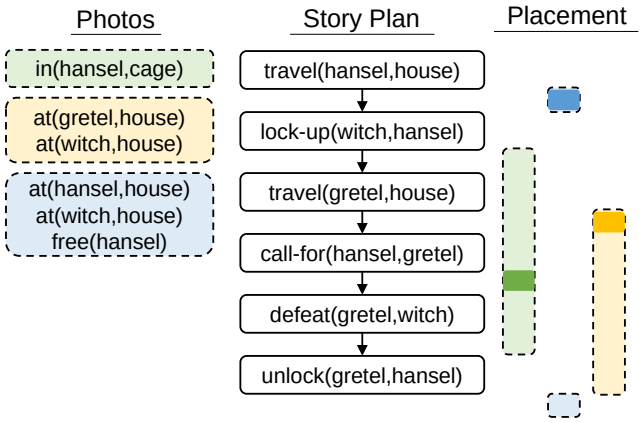


Figure 3: Additional decisions are needed for placing photos. For the three photos on the left, the three bars on the right show their possible placement in the story. As an illustration, we manually picked the best locations for these photos, as indicated by the bright-colored segments.

## Photo Placement

Photos in a high-quality story should be aesthetically pleasing by themselves and also exhibit concord with the story. As discussed earlier, we capture aesthetic quality of individual photos as benefits of optional goals in the ILP problem. We now consider the compatibility between photos and the story, which we use to place photos.

A photo may be placed wherever all of its literals are established, which leaves significant freedom for its ultimate position. Figure 3 illustrates one such example, which considers three photos with different semantic content. A story plan is shown in the middle, where each box represents one action. The three colored bars on

the right show possible positions for the three photos respectively. For example, the blue photo can be shown right after the action `travel(hansel,house)` or right after `unlock(gretel,hansel)`. The bright small boxes show the ideal positions for the three photos.

**Photo-Action Affinity** In order to place photos relative to actions, we develop the metric *photo-action affinity* based on structural properties of the story plan, which we believe to capture how well a photo depicts a story action. The affinity can be positive or negative. In general, a negative affinity indicates the photo and the action should not be placed together. Our metric depends on four features that relate an action $a$ and a discourse goal $g^o$:

$$\text{Support}(a, g^o) = \frac{\left|\text{EFF}(a) \cap g^o\right|}{\left|g^o\right|}$$

$$\text{Conflict}(a, g^o) = \frac{\left|\{l \mid l \in \text{EFF}(a) \wedge \neg l \in g^o\}\right|}{\left|g^o\right|}$$

$$\text{Synergy}(a, g^o) = \frac{\left|\text{PRE}(a) \cap g^o\right|}{\left|\text{PRE}(a) \cup g^o\right|}$$

$$\text{CharSym}(a, g^o) = \frac{\left|\text{CHAR}(a) \cap \text{CHAR}(g^o)\right|}{\left|\text{CHAR}(a) \cup \text{CHAR}(g^o)\right|}$$

where $\text{CHAR}(\cdot)$ denotes the set of story characters involved in an action or a set of literals. In plain English, $\text{Support}(a, g^o)$ measures how many of $g^o$'s literals are established by $a$. $\text{Conflict}(a, g^o)$ measures opposing conditions $a$ sets up against $g^o$, which must then be toggled for achieving $g^o$. $\text{Synergy}(a, g^o)$ measures common preconditions of $a$ and literals in $g^o$, as identical conditions can be achieved simultaneously. CharSym measures story characters shared between the action and the photo of $g^o$. We define photo-action affinity as the linear combination of the four features:

$$\begin{aligned}\text{Affinity}(a, g^o) = \beta_1 \text{Support}(a, g^o) + \beta_2 \text{Conflict}(a, g^o) \\ + \beta_3 \text{Synergy}(a, g^o) + \beta_4 \text{CharSym}(a, g^o)\end{aligned}$$

where $\beta_1, \beta_2, \beta_3$, and $\beta_4$ are linear weights. For the experiment in this paper, we tuned the weights manually, but learning these weights as a linear regression is also straightforward.

**Optimizing Placement** To place photos using the photo-action affinity metric, we treat photo placement as an assignment problem: each photo can be assigned to one action in the story plan, and we seek the assignment that maximizes overall affinity. After a narrative plan is returned by the story planner, we set up an $n \times m$ matrix $\mathcal{M}$ for $n$ actions and $m$ discourse goals, where the element $\mathcal{M}_{ij}$ represents the photo-action affinity between the $j$-th photo and the $i$-th action. If the $j$-th photo does not have its literals true after execution of the $i$-th action, we set that photo-action affinity score to $-1$. After setup, we used the Hungarian algorithm to solve the assignment problem in $O(n^3)$ time.[1] This algorithm requires that all matrix entries are non-negative. Before running the algorithm, we increase all scores by the absolute value of the largest negative entry, and shift them back after running the algorithm.

If, in the returned result, a photo is assigned to an action with a negative affinity, we remove that photo from the generated story. Because of this, some photos may be excluded; we require that every shown photo is accompanied by one action such that it has some accompanying text. It is possible to avoid this issue by tweaking the ILP formulation and letting ILP select photo-action pairs instead of only photos, but we defer that for future work. In this paper, we focus on the affinity-based photo placement strategy and evaluate it in the next section.

[1] Assuming that $n > m$. Otherwise, it runs in $O(m^3)$ time.

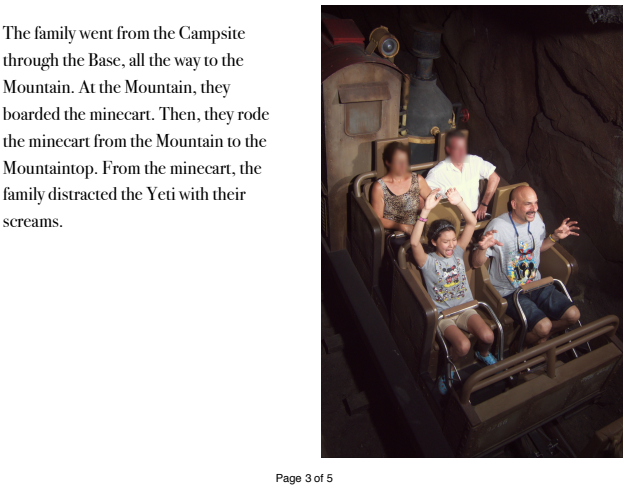


Figure 4: One page in the story booklet generated by the AFFINITY strategy.

Table 1: Weights used for the AFFINITY strategy.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\beta_1$ | 0.25 | $\beta_2$ | -0.5 | $\beta_3$ | 5.0 | $\beta_4$ | 3.0 |

## Evaluation

In this section, we evaluate the photo placement strategy based on the affinity measure with a user study.

**Methodology** For comparison, we created two heuristics for photo placement. The RANDOM strategy places a photo at a random legal location. The EARLY strategy places a photo at the earliest legal location.[2] Weights used for computing photo-action affinity in our affinity-based strategy (AFFINITY) are shown in Table 1.

All photo placement strategies started with the same story plan that involved 16 actions in a Disney-themed domain and 5 photos. One story was created from each placement strategy. Each story was presented to participants as a photo booklet with landscape letter-sized pages. In each booklet, one page contained a maximum of one photo. A photo was always placed on the same page with the action it was paired with. Actions were translated into text using rule-based templates. Figure 4 shows one page from the

[2] We experimented with a LATE heuristic, which always places the photo at the last legal position. Since this heuristic works poorly in practice, we excluded it from this study.

AFFINITY story. Every photo had at least 3 possible slots (avg $= 4.75$) where it could be placed, except for one photo that was always at the end.

We recruited 21 participants, with an age range from 21 to 40, all of whom have at least a bachelor's degree. Each participant read all three stories in randomized ordering. After reading an entire story, participants were asked to provide two five-point Likert-scale ratings for each page. The first scale asked how well each photo was positioned relative to the text on the same page and adjacent pages. The second scale asked how coherent was the juxtaposition of the photo and the text. After reading all three stories, participants were asked to identify the best and worst stories.

**Results** Our results are shown in Table 2 and Table 3. The fourth photo in the RANDOM story does not have accompanying text, so the coherence question does not apply. Our main hypothesis was that photo placement with affinity is better than the other two strategies.

We first analyzed the final story ranking provided by the participants. Our main hypothesis was converted to three directly testable hypotheses: the AFFINITY story is better than the EARLY story; EARLY is better than RANDOM; and AFFINITY is better than RANDOM. Since each is a binary decision, we used a one-tail hypothesis test over the binomial distribution. We reject three null hypotheses at the $p = 0.05$ level, indicating the AFFINITY story is preferred to other baselines.[3]

We then investigated the positioning and coherence Likert scales. We performed Page's trend test (1963) for each of the five photos in order to account for individual photo differences, per condition (Position and Coherence), except for the fourth photo's coherence with text. This is a nonparametric test for within-subject ordering under the alternate hypothesis $Md_{\text{AFFINITY}} > Md_{\text{EARLY}} > Md_{\text{RANDOM}}$, where $Md$ denotes the median rating. In all 9 tests, we reject the null in favor of our alternate at the $p = 0.05$ level, indicating that the AFFINITY story is perceived to have the best photo positioning and is the most coherent.

**Discussion** Although we expected the AFFINITY-based approach to comfortably beat the RANDOM baseline, the comparison between AFFINITY and EARLY was initially less clear: the EARLY heuristic easily beat the RANDOM baseline and received reasonably good ratings, which suggests it often works well and is not a weak baseline.

Therefore, beating the EARLY baseline demonstrates the effectiveness of our approach. This suggests that there is merit to reasoning over the structural relationships between fabula and discourse materials beyond merely how fabula causally enables discourse. Rather, multiple aspects of concord between fabula and discourse can play a role in the perceived quality of the generated artifact. Note that we do not claim the weights in Table 1 are optimal, but rather that there exists a configuration of these features that capture important relationships between fabula and discourse.

[3]The Bonferroni correction does not apply as we need to reject at least two out of three null hypotheses to show our main hypothesis.

Table 2: Results from the ranking of three stories.

| Strategy | Votes | $p$-value |
|---|---|---|
| AFFINITY > EARLY | 16/21 | 0.013 |
| AFFINITY > RANDOM | 20/21 | <0.001 |
| EARLY > RANDOM | 21/21 | <0.001 |

Table 3: Summary of results from Likert-scale ratings for position and coherence. All Page trend tests are significant at the $p = 0.05$ level.

| Criterion | Median Score | Photo Index 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Position | AFFINITY | 5 | 3 | 5 | 5 | 4 |
| | EARLY | 5 | 2 | 2 | 4 | 4 |
| | RANDOM | 2 | 3 | 2 | 1 | 3 |
| | Page's L-stat. | 453 | 365 | 469 | 474 | 475 |
| Coherence | AFFINITY | 5 | 3 | 5 | 5 | 4 |
| | EARLY | 5 | 2 | 2 | 4 | 4 |
| | RANDOM | 2 | 3 | 2 | N/A | 3 |
| | Page's L-stat. | 467 | 437 | 450 | N/A | 453 |

## Conclusions and Future Work

Our work represents a unique departure from the fabula-driven model of story generation. To the best of our knowledge, PLOTSHOT is the first system that can selectively incorporate discourse materials during the process of fabula generation. This capability is due to our novel formulation of the story planning problem, which combines classical planning with oversubscription planning. Our pipeline addresses multiple considerations for generating an illustrated story around user-supplied photos. The evaluation demonstrates that an informed photo placement strategy can beat non-informed baseline techniques. All told, we have demonstrated that complex decision making is needed to ensure the coherence between fabula and discourse materials during story generation.

As a first step toward selective utilization of discourse constraints during fabula generation, we believe this paper opens new research avenues. One such avenue is the development of a unified fabula/discourse search paradigm, wherein discourse can constrain fabula during generation and vice-versa. Another avenue looks at evaluating design trade-offs of our approach. We briefly discussed an alternative to our photo placement strategy that selects action-photo pairs using ILP. Arguably, each strategy represents a different trade-off point between generative flexibility and local coherence of the discourse materials. Exploring each trade-off point's expressive range (Smith and Whitehead 2010) could be insightful.

Reasoning independently about fabula and discourse has proven useful for computationally generating interesting stories. We believe that exploring their interdependencies will bring AI story generation systems to new heights.

## Acknowledgments

We thank Darrin Bentivegna for letting us use his photos, Kyna McIntosh and Moshe Mahler for the artwork.

## References


Aylett, R.; Dias, J.; and Paiva, A. 2006. An affectively driven planner for synthetic characters. In *Proceedings of the 16th International Conference on Automated Planning and Scheduling*, 2–10.

Barot, C.; Potts, C. M.; and Young, R. M. 2015. A tripartite plan-based model of narrative for narrative discourse generation. In *Proceedings of the Joint Workshop on Intelligent Narrative Technologies and Social Believability in Games at the 11th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 2–8.

Blum, A. L., and Furst, M. L. 1997. Fast planning through planning graph analysis. *Artificial Intelligence* 90(1):281–300.

Dai, Q.; Carr, P.; Sigal, L.; and Hoiem, D. 2015. Family member identification from photo collections. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 982–989.

Domshlak, C., and Mirkis, V. 2015. Deterministic over-subscription planning as heuristic search: abstractions and reformulations. *Journal of Artificial Intelligence Research* 52:97–169.

García-Olaya, A.; de la Rosa, T.; and Borrajo, D. 2011. Using the relaxed plan heuristic to select goals in oversubscription planning problems. In *Advances in Artificial Intelligence*. Springer Berlin Heidelberg. 83–85.

Gervás, P. 2009. Computational approaches to storytelling and creativity. *AI Magazine* 30(3):49.

Joshi, D.; Datta, R.; Fedorovskaya, E.; Luong, Q.-T.; Wang, J. Z.; Li, J.; and Luo, J. 2011. Aesthetics and emotions in images. *IEEE Signal Processing Magazine* 28.5:94–115.

Krishna, R.; Zhu, Y.; Groth, O.; Johnson, J.; Hata, K.; Kravitz, J.; Chen, S.; Kalantidis, Y.; Li, L.-J.; Shamma, D. A.; Bernstein, M.; and Fei-Fei, L. 2016. Visual genome: Connecting language and vision using crowdsourced dense image annotations. In *arXiv preprint arxiv:1602.07332*.

Li, B. 2015. *Learning Knowledge to Support Domain-Independent Narrative Intelligence*. Ph.D. dissertation, Georgia Insitute of Technology.

Miller, C. E.; Tucker, A. W.; and Zemlin, R. A. 1960. Integer programming formulation of traveling salesman problems. *Journal of the ACM* 7(4):326–329.

Montfort, N.; Pérez, R.; Harrell, D. F.; and Campana, A. 2013. Slant: A blackboard system to generate plot, figuration, and narrative discourse aspects of stories. In *Proceedings of the 4th International Conference on Computational Creativity*, 168–175.

Morgens, S.-M., and Jhala, A. 2013. Synthetic photographs for learning aesthetic preferences. In *Proceedings of the Workshop on Late-Breaking Developments in the Field of Artificial Intelligence at the 27th AAAI Conference on Artificial Intelligence*, 83–85.

Page, E. B. 1963. Ordered Hypotheses for Multiple Treatments: A Significance Test for Linear Ranks. *Journal of the American Statistical Association* 58(301).

Piacenza, A.; Guerrini, F.; Adami, N.; Leonardi, R.; Porteous, J.; Teutenberg, J.; and Cavazza, M. 2011. Generating story variants with constrained video recombination. In *Proceedings of the 19th ACM International Conference on Multimedia*, 223–232.

Radiano, O.; Graber, Y.; Mahler, M. B.; Sigal, L.; and Shamir, A. in revision. Story albums: Creating fictional stories from personal photograph sets. *Computer Graphics Forum*.

Riedl, M. O. 2009. Incorporating authorial intent into generative narrative systems. In *Proceedings of the AAAI Spring Symposium on Intelligent Narrative Technologies II*.

Robertson, J., and Young, R. M. 2014. Finding schrödinger's gun. In *Proceedings of the 10th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 153–159.

Ronfard, R., and Szilas, N. 2014. Where story and media meet: computer generation of narrative discourse. In *Proceedings of the 5th Workshop on Computational Models of Narrative*, 164–176.

Smith, G., and Whitehead, J. 2010. Analyzing the expressive range of a level generator. In *Proceedings of the 2010 Workshop on Procedural Content Generation in Games at the 5th International Conference on the Foundations of Digital Games*, 4–10. ACM.

Smith, D. E. 2004. Choosing Objectives in Over-Subscription Planning. In *Proceedings of the 14th International Conference on Automated Planning and Scheduling*, 393–401.

Teutenberg, J., and Porteous, J. 2015. Incorporating global and local knowledge in intentional narrative planning. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, 1539–1546.

Tomai, E. 2014. Exploring abductive event binding for opportunistic storytelling. In *Proceedings of the 10th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 181–187.

Ware, S. G., and Young, R. M. 2014. Glaive: A state-space narrative planner supporting intentionality and conflict. In *Proceedings of the 10th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 80–86.

Young, R. M.; Ware, S.; Cassell, B.; and Robertson, J. 2013. Plans and Planning in Narrative Generation: A Review of Plan-Based Approaches to the Generation of Story, Discourse, and Interactivity in Narratives. *Sprache und Datenverarbeitung: International Journal of Language Processing* 37(1–2):67–77.

Young, R. M. 2007. Story and discourse: A bipartite model of narrative generation in virtual worlds. *Interaction Studies* 8(2):177–208.