

# Technical Disclosure Commons

---

Defensive Publications Series

---

19 Nov 2024

## Photo Composition Guidance based on Scene and Subject Analysis

Kshitij Suresh Gaikar

Pushkar Mittal

Follow this and additional works at: [https://www.tdcommons.org/dpubs\\_series](https://www.tdcommons.org/dpubs_series)

---

### Recommended Citation

Gaikar, Kshitij Suresh and Mittal, Pushkar, "Photo Composition Guidance based on Scene and Subject Analysis", Technical Disclosure Commons, (November 19, 2024)

[https://www.tdcommons.org/dpubs\\_series/7573](https://www.tdcommons.org/dpubs_series/7573)



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

## **Photo Composition Guidance based on Scene and Subject Analysis**

### **ABSTRACT**

This disclosure describes techniques to guide a user in the composition of photos captured by a camera device, based on automatic scene analysis and subject framing. A camera device prompts the user to pan the device across a scene surrounding the user to capture a scene photo, e.g., a 360-degree panoramic photo. Additionally, the user captures a subject photo showing an intended subject for a new photo. An artificial intelligence (AI) model is used to analyze the scene photo and the subject photo. The model outputs locations and/or poses for the subject in the scene that makes for an appealing photo. Composition options are presented to the user that highlight suitable locations and/or poses for the subject and provide composition guidance, which the user can follow when capturing the new photo. The described techniques can enable users to capture high quality photos with a device based on real-time recommendations based on established composition techniques.

### **KEYWORDS**

- Photo capture guidance
- Scene analysis
- Subject analysis
- Photo composition
- Photo framing
- Pose guidance
- Subject pose
- Multimodal LLM
- On-device LLM

## BACKGROUND

Users capture photos using camera devices (e.g., smartphones, cameras, or other devices) in a variety of environments and surroundings. However, photo captures, particularly by non-expert users, are frequently hindered by several practical limitations to the visual quality of a captured photograph. For example, photos may have poor composition and framing due to difficulties in identifying pleasing camera angles, subject placement within the frame, and overall scene composition. This may be due to a lack of familiarity with established photographic principles such as the rule of thirds, leading to imbalanced or unappealing photographs.

Furthermore, the visual quality of captured images is highly susceptible to variations in lighting. Harsh or uneven illumination, as well as low-light environments, can result in images that are overexposed, underexposed, or lacking in clarity and detail. In addition, visually poor subject-background relationships may affect the quality of photos. For example, a cluttered or distracting background can diminish the visual impact of a photograph, drawing attention away from the intended subject. The selection of an appropriate background is often complicated by interaction between the background and subject characteristics such as skin tone, proportions, or attire.

## DESCRIPTION

This disclosure describes techniques for providing composition guidance and options to a user when using a camera device to capture a subject in a photographic scene. Practical challenges to capturing high quality photos are overcome by the use of artificial intelligence (AI) techniques that guide users in capturing visually pleasing photographs through intelligent scene analysis, subject framing, and composition guidance, even in complex or challenging

environments. The techniques are implemented with user permission to perform such analysis. Users are provided options to turn off use of the techniques, or to use them selectively for specific photographs.

Per the techniques, an AI model is used to scan and analyze a scene and an intended subject, determine framing for the subject based on established photographic principles, and provide real-time guidance on the camera device to assist the user in capturing a visually pleasing photo. Features include scene analysis in which the capture area (a 360-degree or other panoramic scene surrounding the user) is scanned to assess scene characteristics such as lighting conditions, points of interest, and overall composition. Subject analysis can include analyzing a photo for a subject's pose, body language, attire, and relationship to the background. Based upon the analysis, suggestions are provided to the user to capture a photo having pleasing locations, poses, and framing for the subject within the surrounding scene. These features can substantially reduce the user's difficulty in finding a suitable camera orientation, scene framing, and composition, as well as in coping with challenging lighting conditions and ensuring the subject complements the background in a captured photo.

The described techniques can be implemented on any suitable camera device that can capture and display images, e.g., a portable user device (e.g., a smartphone, wearable device, etc.), desktop computer, laptop computer, virtual reality or augmented reality headset, etc. The AI model is implemented on-device.

### *Photo capture procedure*

An example set of steps for photo capture utilizing assistance from an AI model are described below.

### User arrival and initiation

A user arrives at a desired location to capture a photograph of a scene at the location. The user can launch or access a camera application on a camera device and activate the photo composition function of the app.

### User captures scene

The user can capture the surrounding scene in a scene photo using the camera device. For example, the user may capture a panoramic photo of their surroundings, such as a 360° view or other panoramic view. The camera device can prompt the user to pan the device across the scene to capture it.

During this process, the camera device also may detect (with user permission) relevant metadata, including ambient lighting conditions, prominent subjects within the scene (e.g., persons, animals, etc. above a threshold size), prevailing color schemes, thematic elements (e.g., objects that may indicate the type of scene or event at the scene, such as a monument, park, city street, furniture, birthday cake, basketball, etc.), geographic location of the device or other information, etc.

### Subject selection

The camera device prompts the user to select the intended subject of the desired photograph. For example, the user can use the camera device to take a subject photo of the intended subject within the scene, and the user designates this subject photo for the intended subject. In another example, the user can select a stored photo of the intended subject that was previously captured.

### AI analysis

An AI model processes both the scene photo and the subject photo, as well as any collected metadata, to analyze various elements in each photo. For example, a multimodal large language model (LLM) or other suitable type of AI model can be used. The AI model can be trained or prompted with established photo composition techniques that can be the basis for its recommendations for composition and poses. For the scene photo, the AI model is utilized to assess lighting conditions in the scene, points of interest (e.g., objects in the scene), overall composition (e.g., placement of sky, floor, objects, etc.), and background in the photo. For the subject photo, the AI model analyzes the subject's pose, body language, and attire, and the relationship of the subject to the background of the scene photo. Locations within the scene photo that are best suited to the subject are determined by the AI model. For example, subject-background matching is ensured to provide high visual quality.

### Recommendations and guidance

The camera app provides the AI-generated recommendations and suggestions on the camera device for the user. The guidance can be displayed, provided as audio, or otherwise output by the device. The recommendations may include visual cues and suggestions, highlights of the best location(s) in the scene photo to include the subject, suggestions of suitable poses and aesthetic framing for the subject, etc. For example, the recommendations can include a suggested position for the subject within the scene, highlighting the most visually appealing areas within the 360-degree scene corresponding to the scene photo. Pose and framing suggestions can be based on body language and/or attire of the subject in the subject photo. Suggestions can be based on subject-background color and/or contrast harmony that enable the subject to stand out and complement the surroundings.

In some cases, the recommendation feature may generate and present multiple composition options to the user on the camera device, allowing the user to choose between them. The most visually appealing and technically sound locations and poses for the subject can be highlighted. The suggestions can be made in real time, e.g., displayed on the viewfinder display screen of the camera device over or alongside a displayed portion of the scene as the user points the camera device at various locations of the surroundings that correspond to the analyzed scene photo.

In some cases, the recommendations can include text suggestions including composition advice for a photo, such as: “position the subject person slightly off-center to the right, utilizing the rule of thirds”; “position the subject person centrally, directly in front of the building’s entrance, emphasizing the symmetry of the structure”; or “have subject person stand on the steps leading up to the building, using the leading lines of the steps to draw the viewer’s eye towards the main structure.” The recommendations can also include graphical suggestions, such as a box or outline displayed at a particular location in the scene that is displayed on the viewfinder screen of the camera device.

The recommendations can include pose and framing suggestions, including suggestions of flattering poses and compositions based on body language, attire, and background elements. For example, the recommendations can include suggestions including pose and framing advice for the subject, such as: “use the ornate railing on the left as a natural frame, partially obscuring the background and creating a sense of depth. Capture a portion of the sky for negative space, adding a feeling of openness and tranquility”; or “utilize a full frame approach, filling the image with the subject and the ornate entrance. The door and surrounding decorations act as a natural frame, further highlighting the subject.”

The recommendations can also include real-time guidance on camera positioning and framing that allow the photographer to be included in the photo (e.g., by setting the shutter for delay using a timer, etc.), if the camera app is so instructed by the photographer. A location in the scene is suggested that allows space for the photographer to be included in the photo. This is determined from the outset when framing the photo before capture. For example, the 360-degree scene scanning allows for a comprehensive understanding of the environment and effective identification of camera orientation that can include additional subjects (such as the photographer), even in complex or dynamic scenes.

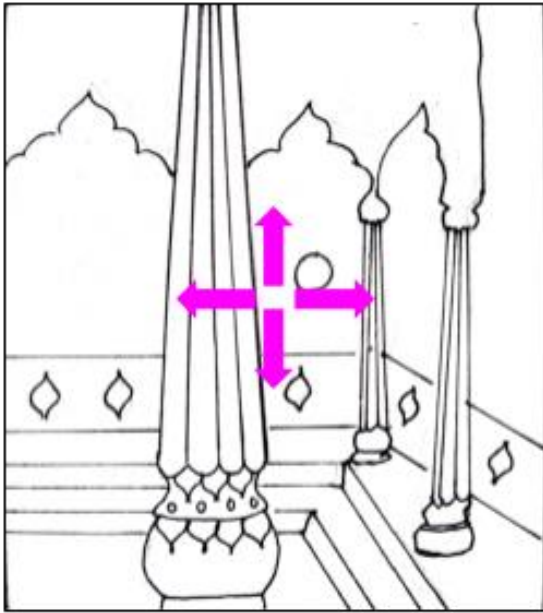
#### User captures photo

Based on the recommendations, the user can direct or position the subject at a particular location in the surrounding scene and pose the subject in a particular way. The user can then capture a photo that includes the subject with a high-quality composition with appropriate lighting, background and subject agreement, and other attributes.

#### *Examples*

Figures 1-4 illustrate example photos taken in a photo capture procedure that incorporates features described above.





**Fig. 1: Example scene photo**

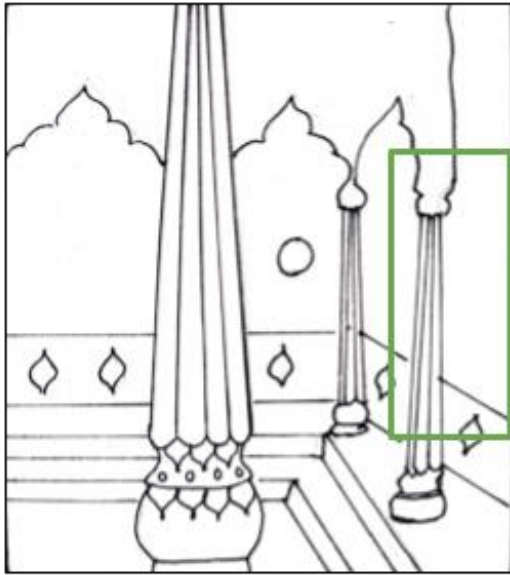


**Fig. 2: Example subject photo**

Fig. 1 shows an example scene photo that has been captured by the user in a particular location using a camera device. This scene photo can be a 360-degree photo (indicated by the arrows) or other type of panoramic photo that can include all or most of the scene surrounding the user at the user's location.

Fig. 2 shows an example subject photo captured by the user with the camera device. For example, the camera device of the user issued instructions to the user to capture the subject in the scene depicted in the scene photo, and in this example a subject person is captured at a location corresponding to a location in the scene photo.

The scene photo and the subject photo are analyzed by an AI model to determine a location within the scene (of Fig. 1) for the subject (the person in Fig. 2) to be positioned for a visually appealing photo. Additionally, the AI model can also determine and suggest a pose and framing based on the scene photo and the subject photo.



**Fig. 3: Example of subject location suggestion**



**Fig. 4: Example final photo**

Fig. 3 shows an example of a recommendation provided by a camera app on a display screen of the camera device, indicating a location within the scene to place the subject, as determined by the AI model. In this example, the suggested location is highlighted by a green rectangle that is displayed over the scene photo on the viewfinder display screen of the camera device. As the user adjusts the position or orientation of the camera device in real time and a changing view of the scene is displayed on the screen, the green rectangle is consistently displayed at this location. Multiple recommendations are provided, and other locations within the scene can be similarly highlighted as the user moves the camera device to view those locations, to indicate other recommended location options to position the subject.

The AI model can alternatively or additionally provide recommendations that are text suggestions for subject placement (e.g., “subject should stand between pillar and fence near corner of fence”). Further, the AI model provides text suggestions as to subject pose and framing, e.g., for the subject to face out toward the horizon, for the city and sky to be framed in

the background of the photo, etc. Fig. 4 shows an example final photo that the user has captured with the camera device, based on the recommendation of Fig. 3 and pose suggestions provided by the camera app.

The described photo composition features provide several advantages. These techniques overcome photo composition challenges by having an AI model automatically analyze a scene and subject holistically and guide a user toward visually appealing compositions, eliminating guesswork and uncertainty. Lighting-related concerns are addressed, where the AI model considers lighting conditions within the scene photo and suggests camera positioning and orientation to ensure the subject is well-lit and the photo is properly exposed. Subject-background harmony is also achieved, where the AI model analyzes the relationship between characteristics of the subject and the background of the scene photo.

Real-time guidance can be provided to the user within a camera app, enabling users to make informed decisions before and during capture of a photo. These features can be integrated seamlessly into an existing camera app, offering a user-friendly and intuitive experience. The guidance features may incorporate 360-degree (or other panoramic) scene scanning and subject analysis, allowing for a comprehensive understanding of the environment and subject characteristics and not only focusing on facial analysis within a frame. With user permission, subject analysis can encompass a wide range of factors such as skin tone, attire, and overall body language. The described real-time guidance can suggest adjustments to the subject's pose or position to further enhance the portrait.

These features can empower users of varying skill levels to take professional-looking photos without extensive knowledge of photography techniques. These features can be provided in a variety of use cases including, for example, portrait photography (suggesting pose, framing,

and background for individual or group portraits), landscape photography (suggesting vantage points and composition for capturing scenic views), selfies (guiding users on angles, poses, and expressions for self-portraits), and object photography (suggesting framing and background choices to highlight the object of interest).

Users are provided with options to grant permissions to and/or to disable described features entirely. The various features of the system are implemented only with user permission to access user information that serves as input to the system (e.g., user images and/or videos, user context information, camera input, user's preferences, etc.). Users are provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information, and if the user is sent content or communications from a server.

## CONCLUSION

This disclosure describes techniques to guide a user in the composition of photos captured by a camera device, based on automatic scene analysis and subject framing. A camera device prompts the user to pan the device across a scene surrounding the user to capture a scene photo, e.g., a 360-degree panoramic photo. Additionally, the user captures a subject photo showing an intended subject for a new photo. An artificial intelligence (AI) model is used to analyze the scene photo and the subject photo. The model outputs locations and/or poses for the subject in the scene that makes for an appealing photo. Composition options are presented to the user that highlight suitable locations and/or poses for the subject and provide composition guidance, which the user can follow when capturing the new photo. The described techniques can enable users to capture high quality photos with a device based on real-time recommendations based on established composition techniques.

## REFERENCES

1. “Gemini Flash - Google DeepMind” available online at <https://deepmind.google/technologies/gemini/flash/> accessed Nov 14, 2024.
2. Donsbach, Aaron Michael, Christopher Breithaupt, Li Zhang, Arushan Rajasekaram, and Navid Shiee. "Systems and methods for providing feedback for artificial intelligence-based image capture devices." U.S. Patent 11,403,509, issued August 2, 2022.
3. “Artificial Intelligence and Photography,” available online at <https://www.mapleleafpictures.com/blogs/artificial-intelligence-and-photography>, accessed Nov 4, 2024.
4. “AI in Photography: How iPhone Uses AI for Better Photos,” available online at <https://www.almabetter.com/bytes/articles/exploring-the-role-of-ai-in-i-phone-photography>, accessed Nov 4, 2024.