1. Introduction

Digital photographs are usually organized by text labeling in order to search and browse a collection. However, labeling many photographs costs users much effort. On the other hand, our target photographs are enhanced with spatial metadata such as geographic coordinates where they were taken and the directions toward where they focused on. These metadata are generated by spatial sensors such as GPS and gyrocompasses, and embedded in headers of image files in Exif format [4], which is a widely used standard for storage of information about cameras, images and shooting conditions within JPEG and TIFF image files. We assume such photographs will be popular. The aim of this paper is sharing text labels among such photographs. A simple method is propagating label data based on geographic proximity of positions where photographs were taken. However, even if two photographs were taken from the same position, they may show different directions, and quite different scenes, and a text label for one photograph may not be appropriate to the other photograph. The reason of the problem is that used geographic coordinates are positions of cameras, and they are not positions of objects taken in photographs and pointed by label texts. To manage the problem, we propagate labels based on whether their pointing objects are seen in photographs. For that purpose, we compute geographic locations of objects pointed by those labels at first, and give those labels to each photograph based on the spatial relationships between geographic locations of labels and field of view of the photograph. We have implemented this process as label sharing system. The system shares label data by the following process:

1. Storing label data submitted by users
2. Computing geographic locations of objects in the real world pointed by stored labels
3. Retrieving a set of label data seen in the photographs, and giving the set to them.

Acquired label data from the system are placed on appropriate positions on photographs, which enable us to access related Web pages by clicking labels in photographs when label data have URLs. The remaining part of this paper is constructed as follows: The target data are defined in Section 2. A propagating method and geocoding method are explained in Section 3 and 4, which are main ideas of this paper. An implemented prototype system is introduced in Section 5, and some concluding remarks and future works are described in Section 6.
2. Data Structure

§ Spatial Photograph

We named target data of our research as *spatial photograph*. It is a photo data enhanced with geographic coordinates of camera’s location, direction, and view angle in the real world. In other words, it is a photograph whose information about from where and toward where it was taken is explicitly recorded. A spatial photograph data is defined as follows:

\[
p = (\text{image}, \text{fov}, L)
\]

: spatial photograph

\[
\text{image}
\]

is a photo image of the photograph. \(\text{fov}\) is a spatial metadata of the photograph. \(L = \{l_1, l_2, l_3, \ldots\}\) is a set of a spatial label data \(l\). A spatial label data is an enhanced text label data for the photograph, and defined in the next sub section. \(\text{fov}\) in spatial photograph is defined as follows:

\[
\text{fov} = (S, \theta_w, \theta_h)
\]

: field of view

\(S = (X_S, Y_S, Z_S)\) is the viewpoint. It is the position of the camera at a time when the photograph was taken. \(V = (X_V, Y_V, Z_V)\) is the view direction. It is the direction where the photograph is focused on. These values are generated by GPS, electric compasses or gyro compasses, and stored in Exif information. Both \(\theta_w\) and \(\theta_h\) are the view angles. They are calculated by using values of focal length and CCD size, which most digital cameras stored in Exif information. Figure 1 shows a geometric model of a spatial photograph. As shown this figure, spatial metadata of a spatial photograph specify Field of View (FOV) of it. In other words, we handle photographs not as point features but as volume features in GIS (Geographic Information System).

§ Spatial Label Data

*Spatial label* data is an enhanced text label data for spatial photographs. In our implementation, when a user attaches a spatial label for a spatial photograph, he or she specifies a certain point on the photograph by clicking on it, and inputs label text. For example, a user labels a photograph showing Tokyo Tower by clicking a point around Tokyo Tower in the photograph, and write “Tokyo Tower.” A user can attach multiple labels to a spatial photograph. We handle only labels for objects associated with locations in the real world, such as “Tokyo Tower”, “Starbucks Cafe”, and so on. Labels such as “My Birthday”, “My Friends” are not appropriate for our framework. A spatial label data can also have geographic coordinates of the location of its pointing object in the real world. In other words, a spatial label data is a kind of POI (Point Of Interest) data, which is a point feature for representing geographic locations of shops, facilities, sightseeing spots, and so on in GIS.

![Figure 1. Spatial Photograph](image1.png)

![Figure 2. Spatial Label](image2.png)
A spatial label data $l$ is defined as follows:

$$ l = \{\text{text}, \url{url}, M, m\} $$

: spatial label

\text{text} is a label test. It is a name of geographic object pointed by the label data such as “Tokyo Tower.” \url{url} is URL on the Web. $M = (X_M, Y_M, Z_M)$ is geographic coordinates of the pointed object. $m = (x_m, y_m)$ is pixel coordinates of the pointed object on a photograph. Null values are permitted for geographic coordinates $M$ or pixel coordinates $m$. Figure 2 shows a geometric model of a spatial photograph.

3. Propagating Spatial Label Data

For propagating spatial label data among spatial photographs, we search for a set of appropriate spatial label data to each photograph. This section therefore describes the method of giving appropriate spatial labels to a spatial photograph and when a user requires spatial labels for the spatial photograph, and that of displaying those labels on appropriate positions on the photograph. We perform the following steps for the process:

1. search for spatial labels whose pointing objects are seen in the spatial photograph
2. compute pixel coordinates of searched labels on the spatial photograph

The details of the above steps (1) and (2) are described in Section 3.1 and 3.2, respectively. Only spatial labels which have geographic coordinates are handled in this process. In other words, we cannot propagate a spatial label if it does not have geographic coordinates. We therefore propose the method for computing geographic coordinates of a spatial label in Section 4.

3.1. Search for Labels in View Angle

We judge whether a pointed object by a label is seen in a spatial photograph by using their spatial relationships. To be specific, a spatial label is judged to be seen in a spatial photograph when the geographic coordinates of the spatial label is contained in the view angle of the spatial photograph. For example, in Figure 3, spatial label $l_1$, $l_2$, $l_3$ are searched as appropriate labels of spatial photograph $p$.

We use the following function “InView” for describing the judgment.

$$ \text{InView}(fov, Q) = \begin{cases} 
\text{true} & \text{if } Q \text{ is contained in } fov \\
\text{false} & \text{if } Q \text{ is not contained in } fov 
\end{cases} $$

(1)

$fov = (S, V, \theta_s, \theta_v)$ is spatial metadata of a spatial photograph and $Q = (X, Y, Z)$ is a geographic coordinates of a point. The returned value of the function InView is computed by the following formula.

$$ \text{arccos} \left( \frac{Q - \text{fov}.S \cdot \text{fov}V}{\|Q - \text{fov}.S\| \cdot \|\text{fov}V\|} \right) \leq \text{fov} \cdot \theta_v $$

(2)

Function InView returns true if the formula is true for given $fov$ and $Q$. By using the function, a set of spatial label data $L_p$ for a spatial photograph $p$ is described as follows.

$$ L_p = \{ l | \text{InView}(p, fov, l, M) = \text{true} \} $$

(3)
3.2. Display Labels on Photographs

In many map applications, spatial data like text labels having geographic coordinates are displayed on a map by using geographic coordinates. Just like it, we display spatial labels also on spatial photographs. If labels have URLs of associated Web pages, photographs become clickable. In other words, those Web pages are mapped to photographs. In order to display spatial labels having only geographic coordinates to spatial photographs, we compute pixel coordinates of them on those spatial photographs.

Figure 4 shows a geometric model between geographic coordinates \( Q(X,Y,Z) \) and image coordinates \( q(x,y) \). A relationship between them is expressed by so-called colinearity equations [5].

\[
x = -a_4(X - X_s) + a_5(Y - Y_s) + a_6(Z - Z_s) - c(X - X_c) + a_5(Y - Y_c) + a_6(Z - Z_c)
\]

\[
y = -a_4(X - X_s) + a_5(Y - Y_s) + a_6(Z - Z_s) - c(X - X_c) + a_5(Y - Y_c) + a_6(Z - Z_c)
\]

\( S = (X_s,Y_s,Z_s) \) expresses coordinates of the viewpoint. \( \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \) are elements of a rotation matrix \( R \), which rotates the view direction to the Z-axis of the absolute coordinate system. They are computed by using values of the view direction.

\[
\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} = R
\]

\( c \) is a focal length, which is computed by using the values of the view angle \( \theta_w \) and the pixel width \( w \) of the photo image:

\[
c = \frac{w}{2 \tan \frac{\theta_w}{2}}
\]

For spatial photographs, \( (x,y) \) in expression (4), (5) are computed when \( (X,Y,Z) \) are given because \( \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \), \( (X_s,Y_s,Z_s) \) and \( c \) are known. By substituting geographic location of a label \( M = (X_u,Y_u,Z_u) \) for \( (X,Y,Z) \), pixel coordinates of a label \( m = (x_m,y_m) \) are therefore given as \( (x,y) \).
Figure 5 shows an example of a photograph mapped multiple spatial labels. We can display like this figure when all substituted values to (4), (5) (i.e. photograph’s viewpoint, view direction and label’s geographic coordinates) are three-dimensional data. When one of those values are two-dimensional data, labels are displayed like Figure 6 since only pixel coordinate \( x \) is computed and \( y \) is not computed.

4. Geocoding Spatial Label Data

This section describes a method for computing geographic locations of spatial labels. By using the proposed method in Section 3, we can propagate only spatial labels which have geographic locations. However, we assume that a user creates spatial label data by clicking and specifying positions on spatial photographs as described in Section 2, and spatial labels are not attached geographic locations in the process, and we cannot apply the method in Section 3 to them. Though in our implementation, a map interface is available and a user can explicitly specify geographic locations by clicking on the map, it cost a user much effort to specify geographic locations of all spatial labels on map. We therefore propose a method for computing geographic locations of spatial labels using spatial metadata for spatial photographs and spatial labels. In our method, a geographic location of a spatial label is computed when multiple labels for different spatial photographs have the same label text as shown in Figure 7.

We explain the method in assuming exploited spatial data (viewpoints and view directions of spatial photographs) are three-dimensional data. If one of them is two-dimensional data without altitude value, computed geographic coordinates are also two-dimensional data without altitude value. At first, we explain the method using two labels with the same text in Section 4.1, and then the method using three or more labels in Section 4.2.

4.1. Label Geocoding using Two Labels

Figure 4 shows spatial relationships between pixel coordinates of a certain point \( q(x, y) \) on a spatial photograph and its geographic location \( Q(X, Y, Z) \). As shown in this figure, when pixel coordinates of a certain point \( q \) on a spatial photograph is given, the direction from the photo’s viewpoint toward the point’s geographic location \( Q \) is computed. About a spatial label placed on a spatial photograph, we can therefore know the direction from the photo’s viewpoint toward the label’s geographic location. In other words, a half line from the photo’s viewpoint toward the spatial label’s geographic location can be specified. We exploit this half line for computing the label’s geographic location. If two labels placed on different spatial photographs point to the same geographic object, that is, have the same label text, we compute geographic coordinates of those labels as the crossing point of two half lines from those photos’ viewpoint toward the label’s geographic location. This method is based on the assumption that if two labels point to the same object in the real world, their geographic location is the same point. The process is explained in detail below. We use following values based on the definition in Section 2.

\[
\begin{align*}
    l_1 &= (text_1, url_1, M_1, m_1) \quad : \text{spatial label} \\
    l_2 &= (text_2, url_2, M_2, m_2) \quad : \text{spatial label} \\
    p_i &= (image_i, fov_i, L_i), \quad l_i \in L_i \quad : \text{spatial photograph} \\
    p_j &= (image_j, fov_j, L_j), \quad l_j \in L_j \quad : \text{spatial photograph}
\end{align*}
\]
\[ f_{ov_i} = (S_i, V_i, \theta_{ai}, \theta_{bi}) \] : field of view
\[ f_{ov_j} = (S_j, V_j, \theta_{aj}, \theta_{bj}) \] : field of view

\( l_i \) and \( l_j \) are spatial labels for the different spatial photographs \( p_i \) and \( p_j \). Only labels’ geographic locations \( M_i \) and \( M_j \) are unknown in these values. The target of this process is computing \( g_i (= g_j) \). In our implementation, two labels are judged to point to the same object in the real world if their label texts are the same. In this case, the label texts of \( l_i \) and \( l_j \) are the same:

\[ text_i = text_j \] (8)

Figure 8 shows spatial relationships among those values. \( V_{m1}, V_{m2} \) in this figure are defined later.

We use the following function “Direction”, which uses a spatial metadata for a spatial photograph \( f_{ov} \) and pixel coordinates of a point on the photograph \( a(x, y) \), and returns a direction vector \( V_a \) along the line from photo’s viewpoint toward the geographic location of point \( a(x, y) \).

\[ V_a = \text{Direction}(f_{ov}, a) \] (9)

We explain the function “Direction” specifically by using Figure 4. As is described in Section 3, a relationship between pixel coordinates of a point on a photograph and its geographic location in the real world is expressed by colinearity equations (4), (5) as described in Section 3.2. They can be transformed as follows.

\[ X - X_s = (Z - Z_s) \frac{a_2 x + a_4 y - a_6 c}{a_3 x + a_5 y - a_7 c} \] (10)
\[ Y - Y_s = (Z - Z_s) \frac{a_2 x + a_4 y - a_6 c}{a_3 x + a_5 y - a_7 c} \] (11)

We define the functions \( f_1(x, y), f_2(x, y), f_3(x, y) \) as follows.

\[ \frac{a_2 x + a_4 y - a_6 c}{a_3 x + a_5 y - a_7 c} = f_1(x, y) \] (12)
\[ \frac{a_2 x + a_4 y - a_6 c}{a_3 x + a_5 y - a_7 c} = f_2(x, y) \] (13)
\[ Z - Z_s = f_3(x, y) \] (14)

By using these functions, (12) and (13) are transformed as follows.

\[ X - X_s = f_1(x, y) f_3(x, y) \] (15)
\[ Y - Y_s = f_2(x, y) f_3(x, y) \] (16)

By substituting label’s photo location \( m = (x_m, y_m) \) for \( q = (x, y) \) in (14), (15) and (16), direction vector \( V_a \) from the viewpoint toward the geographic location of the pointed object by the label is computed as follows.

\[ V_a = \begin{pmatrix} X - X_s \\ Y - Y_s \\ Z - Z_s \end{pmatrix} = \begin{pmatrix} f_1(x_m, y_m) & f_3(x_m, y_m) \\ f_2(x_m, y_m) & f_3(x_m, y_m) \end{pmatrix} \] (17)

\[ |V_a| = 1 \] (18)
By using the function “Direction”, Direction vectors from the viewpoint of photo $p_i$ toward geographic location of label $l_i$ and that from the viewpoint of photo $p_j$ toward geographic location of label $l_j$ are represented as follows.

$$V_{l_i} = \text{Direction}(p_i, \text{fov}_i, l_i, m_i)$$  \hspace{1cm} (19)$$

$$V_{l_j} = \text{Direction}(p_j, \text{fov}_j, l_j, m_j)$$  \hspace{1cm} (20)$$

Label’s geographic location $M_1$ and $M_2$ is represented as points on half lines from viewpoints $S_i$ along direction $V_{l_i}$ and that from $S_j$ along $V_{l_j}$ as follows:

$$M_1 = S_i + k_1 V_{l_i}$$  \hspace{1cm} (21)$$

$$M_2 = S_j + k_2 V_{l_j}$$  \hspace{1cm} (22)$$

$k_1$ and $k_2$ are real numbers. We compute $M_1$ and $M_2$ as crossing point of these half lines.

$$M_1 = M_2$$  \hspace{1cm} (23)$$

In other words, $g$ is computed by solving the simultaneous equations (21), (22) and (23). However, it is unrealistic that two half lines are crossing in three-dimensional space. That means the simultaneous equations cannot be solved. We therefore solve the following equations (24), (25) and (26) instead of (21), (22) and (23).

$$M'_1 = S_i + k'_1 V_{l_i}$$  \hspace{1cm} (24)$$

$$M'_2 = S_j + k'_2 V_{l_j}$$  \hspace{1cm} (25)$$

$$M_1 = M_2 = \begin{pmatrix} M'_1 : X \\ M'_1 : Y \\ M'_1 : Z + M'_2 : Z \end{pmatrix} = \begin{pmatrix} M'_2 : X \\ M'_2 : Y \\ M'_2 : Z + M'_1 : Z \end{pmatrix} \times \frac{1}{2}$$  \hspace{1cm} (26)$$

$M'_1$ and $M'_2$ are points on half lines from viewpoints $S_i$ along direction $V_{l_i}$ and that from $S_j$ along $V_{l_j}$, and their x and y coordinates are respectively equal. $M_1$, $M_2$ are computed as the same point, which is the middle point of $M'_1$ and $M'_2$.

4.2. Label Geocoding using Multiple Labels

When three or more labels for different spatial photographs have the same label text as shown in Figure 9, half lines from viewpoints toward label’s geographic location do not cross at the same point. However, we compute the geographic location of those labels as a point. At first, we compute all crossing points of all sets of two half lines by the method in Section 4.1, and compute the geographic location of the label as the middle point of those crossing points. To be specific, we name label data point to the same object as $l_1, l_2, l_3, \ldots$. They all have the same label text.

$$l_{i,\text{text}} = l_{j,\text{text}} \quad (1 \leq i, j \leq n, i \neq j)$$  \hspace{1cm} (27)$$

By using $M_i$, which is a geographic location computed from label $l_i$ and $l_j$, geographic location for all the labels $l.M = l_1.M = l_2.M = \ldots$ is computed as follows.

$$l.M = \sum_{i,j=1}^{n} \frac{M_i \cdot M_j}{C_i}$$  \hspace{1cm} (28)$$

Figure 9. Label Geocoding using the Same Label on Multiple Photos
5. Prototype System

Figure 10 is a graphical user interface of our prototype system.

It has visual interfaces for a map, thumbnails of photographs, a selected photograph, and a list of spatial labels. The system realizes the following function:

(a) reading and writing spatial metadata of spatial photographs
   Spatial metadata is embedded in each image file in Exif format.
(b) visualizing spatial photographs on map as photo vectors
   As shown in the map in Figure 10, each spatial photograph is visualized as an arrow named photo vector, which is a vector from its viewpoint along its view direction. A user can select spatial photographs by photo vectors on a map.
(c) adding, removing, and displaying spatial labels on maps and spatial photographs
   Spatial labels have geographic coordinates or pixel coordinates on specific photographs, or both, and they are accordingly displayed on maps or photographs or both.
(d) search for photographs by label texts
   The system handles a simple text search for spatial photographs by label texts, and search terms can be set by clicking one of spatial labels on a map or a spatial photograph. A user can therefore jump from a photo to another sequentially by successively clicking text labels on them.
(e) giving a set of spatial label to a selected spatial photograph
   When a user requires spatial labels for a spatial photograph, Appropriate spatial labels are searched from those which have geographic coordinates by the method described in Section 3.1, and pixel coordinates on the photograph are computed by the method described in 3.2.
(f) geocoding spatial labels
   When a user adds spatial labels having the same label text to two or more spatial photographs, Geographic coordinate of the pointed object by those labels are computed by the method described in Section 4.
(g) propagating a spatial label data among all spatial photographs
   For propagating a selected label, the system searches for photographs in which the selected spatial label is contained in their view angles by the method described in Section 3. In Figure 11, the boxed spatial label on the map is
propagated among photographs whose corresponding photo vectors on the map and thumbnails are displayed. Photographs in this figure are sets of panorama photographs took mechanically by a rotating camera for checking the propagation visually.

Figure 11. Propagating Spatial Labels over Relevant Spatial Photos

6. Conclusion and Future Work
We proposed a new framework for sharing text labels among photographs which have spatial metadata such as geographic coordinates of camera’s positions and directions, and also implemented a prototype system based on the framework. In our framework, label data for spatial photographs are shared by computing geographic location of them. In other words, it is a framework improving reusability of label data by transforming them into a kind of spatial data. As future works, we plan to develop the proposed method to share photo labels among many users through the Web. For that purpose, we will improve the method for deciding a set of labels point to the same object in the real world, which is now processed as simple text matching in this paper. We also plan to exploit values of distances to focused positions of photographs, which are used for automatic focus and stored in Exif information. Though they are not precise values, they can be used in deciding search area of spatial labels contained in field of view of spatial photographs. We also think there is a rough correspondence between the distances and contents of photographs, for example, a photograph contains a portrait of one person, a portrait of a group of people, architecture, a landscape, and so on.

References