EXPLOITING COLOR FOR EDGE EXTRACTION AND LINE SEGMENT STEREO MATCHING IN HIGH-RESOLUTION AERIAL IMAGERY

Stephan SCHOLZE, Theo MOONS, Frank ADE, Luc VAN GOOL
Swiss Federal Institute of Technology
Communication Technology Lab
Computer Vision Group
{scholze, ade, vangool}@vision.ee.ethz.ch
Theo.Moons@esat.kuleuven.ac.be

Working Group III/3

KEY WORDS: Image matching, Multi-spectral data, Automation, Edge extraction, Buildings.

ABSTRACT

This paper investigates into the added value of color information for edge extraction and straight edge segment matching between stereo views. For edge extraction in color images different methods proposed in the literature are evaluated and compared, paying special attention to significance and completeness of the obtained edge-map. To find related edge segment pairs in different views, we apply an odd-man-out scheme: starting with all geometrically possible pairs we first rule out pairs, for which the chromatic information provided by the regions flanking the edge segments is incompatible. To further restrict the number of pairs we compute a chromatic similarity measure based on cross-correlation in the color bands. Both steps result in a significant reduction of candidate pairs, yet no correct pairs get lost. A main application of our technique is for automatic 3D building reconstruction from high resolution aerial images.

1 INTRODUCTION

The automatic reconstruction of 3D city models from high resolution aerial images is still an area of active research. The currently available color images of dense urban scenes have nowadays a resolution of $10 \times 10\ cm^2$ on the ground or better. This huge amount of data with its immanent information is on one hand the key to better solve the correspondence problem, on the other hand it holds enormous complexity. This complexity might be the reason, why color information was not used to its full extent until now.

Our strategy to sytematically reduce the complexity of the problem is based on an iterative scheme. After each reduction step a more sophisticated elimination procedure can be applied, since the number of remaining combinations gets smaller after each step. Since the vast majority of man-made objects, especially building roofs, are of polygonal shape with straight lines delimiting the roof edges, we find it natural to base our method on straight line segments. Thus our method could be labeled feature-based. Moreover, it is fully based on color information; at no time we refer to a grey-level image as data source. The main contributions of our work are: edge detection in the color images and the use of chromatic information for the stereo matching algorithm.

For edge extraction in color images we have evaluated and compared two different methods proposed in the literature. The key problem is the fusion of the gradient information from the individual RGB-bands. The maximum of the norms of the gradients in the three bands turns out to be a good compromise between computational simplicity and completeness of the edge map (Section 2).

For 3D building reconstruction straight line segments at roof edges need to be matched between overlapping views. This task is not straightforward. First, the edge detection step will always produce incorrect results to some degree. True image features might be detected only partially, continuous contours in the image might get fragmented, even cases where important image features are not detected at all can happen in real world applications. Second, due to the weak epipolar constraint, we will find a huge number of possible pairings between extracted edges in different views. A calculation of all geometrically possible 3D line reconstructions would therefore yield mostly futile results. To overcome the geometric ambiguities, we take into account the color of regions flanking the extracted line segments. By comparing the integrated color distribution in the flanking regions of putative pairs, we determine a *statistical chromatic similarity measure*. This step reduces the possible combinations to typically 30 percent, before calculating the 3D reconstruction. For the remaining pairs we calculate the 3D information. A further significant reduction to about 20 percent is now achieved by computing a *cross-correlation based chromatic similarity measure*, exploiting the pixel-wise one-to-one correspondence induced by epipolar geometry (Section 3).

2 EDGE DETECTION IN COLOR IMAGES

Edge detection in color images is not a widely studied subject. In most practical applications, edge extraction is performed on the corresponding grey-level intensity image using 'standard' edge-detectors such as the Canny-operator (Canny, 1986) or adaptions as the Deriche-operator (Deriche, 1987). These operators are based on the norm of the intensity gradient in each pixel. Denoting the smoothed grey-level intensity at pixel (x, y) with $i(x, y)$ one would formally compute

$$\|\nabla i(x,y)\| = \sqrt{\left(\frac{\partial i}{\partial x}\right)^2 + \left(\frac{\partial i}{\partial y}\right)^2} \tag{1}$$

In case of a color image with red, green and blue bands, the image can be represented as a vector valued function $I \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$ where $\mathbf{I}(x, y) = (r(x, y), g(x, y), b(x, y))$ represents the RGB-values of the pixel (x, y). We compared the following combinations of the derivatives in the individual bands for replacing $\|\nabla i(x, y)\|$.

2.1 The Maximum

We consider the vector valued function $\mathbf{I}(x, y) = (r(x, y), g(x, y), b(x, y))$, representing the multispectral image at pixel (x, y). To replace the norm of the intensity gradient $\|\nabla i(x, y)\|$ from the single band case, we choose the maximum of the norm of the gradients in the individual color bands at each pixel position:

$$\|\nabla i(x,y)\| \rightarrow \max\left\{\|\nabla r(x,y)\|, \|\nabla g(x,y)\|, \|\nabla b(x,y)\|\right\} \tag{2}$$

2.2 The Spatial Gradient

In case of a scalar field the direction and magnitude of its strongest change are given by the gradient of the field. This idea can be extended to vector fields (Lee, 1991) and was recently used for edge detection in color images (Zafiropoulos and Schenk, 1999). Again we consider the vector valued image function $\mathbf{I}(x, y) = (r(x, y), g(x, y), b(x, y))$ and an arbitrary direction $\mathbf{n} = (\cos\phi, \sin\phi)$ which is defined by the angle ϕ in the image plane. The objective is to find the direction of strongest change in $\mathbf{I}(x, y)$ at point (x, y).

First we compute the directional derivative of $\mathbf{I}(x, y)$ with respect to $\mathbf{n}$:

$$\frac{\partial \mathbf{I}}{\partial \mathbf{n}} = \begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial b}{\partial x} & \frac{\partial b}{\partial y} \end{pmatrix} \mathbf{n} = \mathbf{J}\mathbf{n} \tag{3}$$

which is equivalent with forming the scalar product of the Jacobian $\mathbf{J}$ of the image function $\mathbf{I}$ with $\mathbf{n}$. As measure of magnitude of change of $\mathbf{I}(x, y)$ as a function of $\mathbf{n}$ one usually chooses the square of the norm of $\mathbf{J}\mathbf{n}$:

$$l^2 = \|\mathbf{J}\mathbf{n}\|^2 = \mathbf{n}^T\mathbf{J}^T\mathbf{J}\mathbf{n} \tag{4}$$

With this choice, note that the $\mathbf{n}^T\mathbf{J}^T\mathbf{J}\mathbf{n}$ is equivalent with the Rayleigh-quotient of $\mathbf{D} = \mathbf{J}^T\mathbf{J}$. Thus the direction of the largest change of $\mathbf{I}(x, y)$ at (x, y) is given by $\max(l^2)$ which is in turn given by the largest eigenvector of $\mathbf{D}$. Since $\mathbf{D}$ is real, the eigenvalues of $\mathbf{D}$ are the squares of the singular values of $\mathbf{J}$. Clearly, computing the spatial gradient is computationally expensive, since we have to perform a singular value decomposition at each pixel position.

2.3 Comparison of both combination schemes

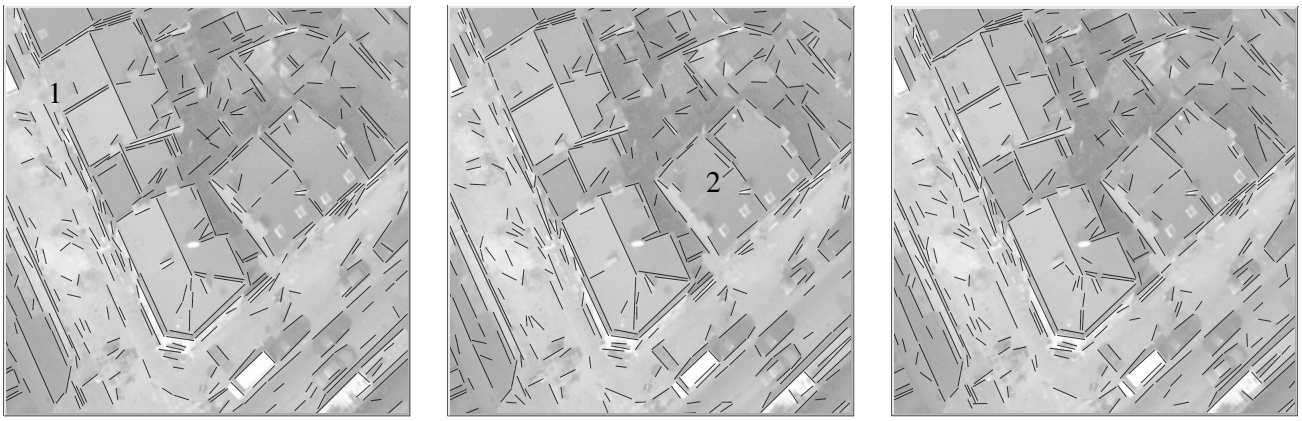
After edge detection and edge linking, straight line segments are fitted to the edges. A very tight threshold is used for line fitting so that curves are not piecewise linearly approximated. Moreover, only line segments above a minimum length (15 pixels) will be considered for matching. For computational efficiency during further processing steps, the line segments are stored in an R-Tree (Guttman, 1981) data-structure. This allows efficient queries for adjacent lines, which will be exploited later. Additionally the straight lines are also stored in raster format, i.e. a pixel indexes a line segment if there is one at that location.

Since we want to investigate polygonal shapes, only the fitted straight line segments will be used all through the rest of the paper. We will not refer to the underlying unfitted data. Thus, in the following the terms 'edge' and 'line segment' will be used interchangeably, both referring to the fitted straight line segments.

In Figure (1) the results of both edge detection schemes are compared with the result obtained by performing the edge detection only on the grey-level intensity image. At first glance, the three methods perform similarly well. The number of

extracted straight line segments in the 500×500-pixel picture with a minimum length of 15 pixels is about 300 in each case. The average length of the extracted segments is about 26 pixels. More particular, the edge detection performed on the grey-level image lacks to detect some important image features, e.g. the roof border denoted with '1' in Figure (1, a). The method based on the maximum of gradients of the three color bands shows the tendency to produce longer contiguous line segments, better capturing the underlying image data. Remarkably it fails to detect some important boundaries, e.g. the roof ridge marked as '2' in Figure (1, b). The method based on the vector gradient produces results similar to those of the maximum method. The extra computational load does not pay off, it seems.

Taking into account the completeness of the obtained edge-map and the computational simplicity, we advocate using the maximum of the norm of the gradients in the single bands, especially if a single edge detection method is to be used. In cases where higher computational load is tolerable, it is possible to overcome the individual failures of each method by additionally applying a simple fusion step: starting with the edges obtained via one method, one iterates through the edge-maps obtained by the other methods and adds isolated edges which are not already present in the initial edge-map.



(a) (b) (c)

Figure 1: Edge detection results for the same image using (a) intensity gradient of grey-level image, (b) maximum of gradients in color bands and (c) spatial gradient. The number of detected line segments with a minimum length of 15 pixels after merging collinear lines was 338 / 317 / 343; the average length of the segments was 26.2 / 26.7 / 26.1 pixels respectively.

3 LINE SEGMENT STEREO MATCHING

Suppose line segments have been determined in both images of a given stereo pair by means of a method described in the previous section. Given a line segment in one image, the goal is to find the corresponding line segment (if available) in the second image. A first step in this direction is to select those line segments from the second view, which possibly can match the given line segment in the first view, thus reducing the search space. Two types of criteria are used for selecting these match candidates: *geometric* constraints based on the geometric relations between the images, and *chromatic* constraints based on the comparison of color properties between flanking regions of (possibly) corresponding line segments.

3.1 The geometric constraints

- Epipolar strip constraint
  Given a line segment $\mathbf{l}$ in the first view, the epipolar relation between two images of a stereo pair yields a weak constraint on the matching line segment in the second view. As depicted in Figure (2), the two endpoints $\mathbf{p_1}$ and $\mathbf{p_2}$ of the given line segment $\mathbf{l}$ define two epipolar lines in the second view ($\mathbf{l}'_{e,\mathbf{p_1}}, \mathbf{l}'_{e,\mathbf{p_2}}$), on which the corresponding points must be found. It is clear that the line segment $\mathbf{l}'$, corresponding to $\mathbf{l}$ in the second view, must have a nonempty intersection with the image region between the two epipolar lines called the epipolar strip. Hence only the line segments that are contained at least partially in this epipolar strip in the second view can be candidate matches for $\mathbf{l}$.

- Further geometric constraints
  Due to image acquisition, the relative orientation of the two images is known. This restricts possible candidates $\mathbf{l}'$ to have less than a maximum angle deviation from the line segment $\mathbf{l}$ in the first image.

  If we have some indication about the maximum building height in the investigated area, again, due to known external camera calibration, we can restrict the region in the second view, where possible match candidates can be found. Still there is a huge number of candidates and no further geometrical restrictions can be made.

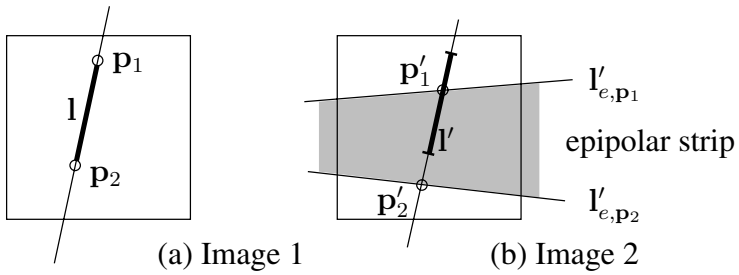


Figure 2: Epipolar constraint: only the line segments $\mathbf{l}'$ that are contained at least partially in the epipolar strip can be candidate matches for $\mathbf{l}$. See description in text.

3.2 The chromatic constraints

In order to disambiguate line segment matching between views, it proves to by very powerful to take into account information about the neighborhood of the line segments under consideration. In (Schmid and Zisserman, 1997) the *intensity* neighborhood of line segments is exploited to determine the optimal match candidate by computing a similarity measure based on cross correlation in the grey-level images. Special attention is payed to determining the correct shape of the correlation window, taking into account projective distortions. The idea to assign *chromatic* region attributes to line segments was used in (Henricsson, 1998) for 2D grouping of line segments and extraction of 2D enclosures. The regions were basically rectangular, although they were adjusted in order not to extend over other line segments in the image.

We develop the above ideas further, by using the chromatic information to rule out wrong match candidates in a twofold way: First, adaptive flanking regions are defined for each side of an oriented line segment and robust estimators are used to characterize the integrated color distribution in each region. For each putative pair we perform a χ^2 test to determine if the mean color vector and the covariance matrix of the flanking regions are compatible, at least for one side. After reducing the complexity of the problem, the remaining candidates are examined in more detail by computing a cross-correlation based similarity measure in the individual color bands, exploiting the pixel-wise one-to-one correspondence induced by epipolar geometry. To capture the underlying geometry, the correlation-mask will be delineated through epipolar lines, as described in Section (3.4).

3.2.1 Construction of adaptive flanking regions The flanking regions are defined by translating the given line segment in the direction normal to it in the image. The rigid extent of this rectangular region is characterized by two parameters: *offset* and *width*. The *offset* quantifies a displacement of the inner region boundary in direction of the normal, away from the line segment. The offset is required to reduce color blurring effects occurring at and around the actual contour. The *width* of the region must be chosen to ensure that the region contains enough data samples (pixels) to allow a robust statistical analysis of the region properties. We use an offset of one pixel and a width of ten pixels.

Technically the regions are obtained via an affine transformation which maps the image coordinates into a coordinate system, where the flanking region is aligned with the coordinate axes. This transformation allows us to introduce a scaling which will be exploited in Section (3.4). To avoid the region from extending over neighboring contours we use the simple morphological operator depicted in Figure (3, b). The operator is defined as follows: include the pixel at position (i, j) in the flanking region, if all three shaded pixels already belong to the region and no neighboring contour is at position $(i, j + 1)$. Thus the region approaches neighboring contours only up to an *offset* of one pixel to reduce color blurring effects as mentioned above.

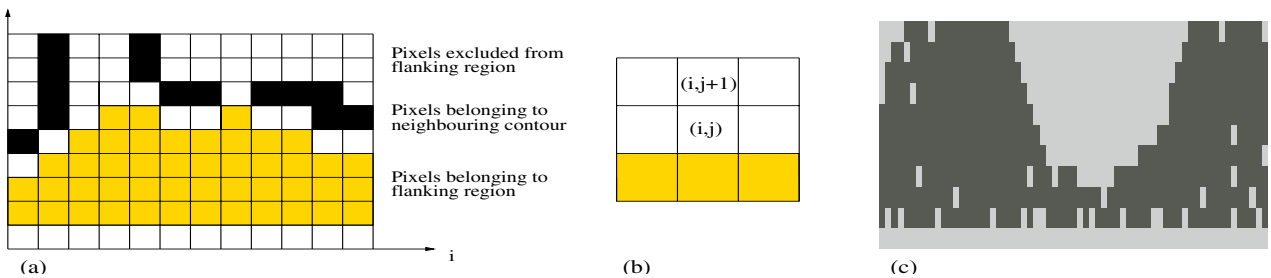


Figure 3: (a): Schematic representation of resulting adaptive flanking region. (b): The morphological 3×3 mask used to grow the region. (c): Example of a flanking region. Pixels belonging to the region are depicted dark. The missing pixels inside the region are chromatic outliers, see Section (3.2.2) for details.

3.2.2 Computing chromatic attributes To each of the flanking regions, we want to assign attributes which represent the chromatic properties of the pixels in the region. Small differences in illumination conditions between images taken in

different flight lines and slightly different operating conditions while scanning the pictures may cause significant changes in the RGB-values of pixels induced from the same roof surface. To overcome this problem it was proposed in (Henricsson, 1996) to use the CIE$(L^*a^*b^*)$ color space for computing the chromatic attributes of the flanking regions.

Since differences in chroma may only be derived if the stimuli have the same lightness, in a first step pixels with outlier L^*-values have to be removed from the region. To identify and discard the L^*-outlier pixels we apply a Minimum Volume Ellipsoid (MVE) estimator (Rousseeuw and Leroy, 1987). It is defined as the ellipsoid of minimal volume that covers at least half of the data points. For increasing sample size, the breakdown point of the MVE estimator converges to 50 % (Rousseeuw, 1985), yielding robust estimates. Next, we compute the MVE estimates for the mean vector and covariance matrix of the a^*-b^*-distribution, only considering those pixels which are belonging to the adaptive region and also are L^*-inliers. The obtained values robustly represent the chromatic properties of the flanking regions on both sides of a given line segment.

3.3 Forming initial match hypotheses

For a given line segment $\mathbf{l}$ in the first view, the preliminary set of possible correspondences is given by all geometrically possible line segments $\{\mathbf{l}'\}$ in the second view. To prune this set further we exploit the chromatic attributes derived in the previous step. We will only keep correspondence candidates, which have similar mean color vector and covariance matrix in at least one flanking region. We do *not* require chromatic similarity on *both* sides of the line segments in order to be robust against frequent cases of occlusion.

Let us now define our tests for similar mean color vector and covariance matrix. Both are applications of the likelihood ratio test for multivariate distributions (Krzanowski, 1993). The color distribution in the flanking regions associated with a straight line segment has a two-variate population in the variables (a^*, b^*). Under the assumption of normality, the estimates for mean μ and covariance matrix $\mathbf{\Sigma}$ are the sample mean $\bar{\mathbf{x}}$ and the sample covariance matrix $\mathbf{S}$.

- Testing the sample mean $\bar{\mathbf{x}}$.
  Consider the sample mean color vector $\bar{\mathbf{x}}_{ab} = (\bar{a^*}, \bar{b^*})$ of a given flanking region in the first view and its considered counterpart $\bar{\mathbf{x}}'_{ab}$ in the second view. Here, $\bar{a^*}$ and $\bar{b^*}$ denote the robustly estimated mean values of the color distribution in the flanking regions in the CIE$(L^*a^*b^*)$ color space. We want to prove the null hypothesis: $H_0 : \bar{\mathbf{x}}'_{ab} = \bar{\mathbf{x}}_{ab}$ against the alternative $H_1 : \bar{\mathbf{x}}'_{ab} \neq \bar{\mathbf{x}}_{ab}$. The likelihood ratio test statistics leads to:

$$t_{\bar{\mathbf{x}}} = n\,(\bar{\mathbf{x}}_{ab} - \bar{\mathbf{x}}'_{ab})\,\mathbf{S}^{-1}\,(\bar{\mathbf{x}}_{ab} - \bar{\mathbf{x}}'_{ab}) \tag{5}$$

  with $\mathbf{S}$ denoting the sample covariance matrix of the a^*-b^*-distribution in the first view and n being the number of data-points (pixels). If H_0 is true, then $t_{\bar{x}}$ has a χ^2_p distribution with a number of degrees of freedom $p = 2$. See below for choosing the appropriate significance level α of the test.

- Testing the covariance matrix $\mathbf{S}$.
  Let us denote the sample covariance matrix of a given flanking region in the first view as $\mathbf{S}$ and its counterpart in the second view $\mathbf{S}'$. Analogous to testing the mean, we want to prove the null hypothesis: $H_0 : \mathbf{S}' = \mathbf{S}$ against the alternative $H_1 : \mathbf{S}' \neq \mathbf{S}$. Again the likelihood ratio test statistics leads to:

$$t_{\mathbf{S}} = n\ \mathrm{trace}(\mathbf{U}) - n \ln |\mathbf{U}| - np \tag{6}$$

  with $\mathbf{U} = \mathbf{S'}^{-1}\mathbf{S}$, n being the number of pixels and p the number of degrees of freedom. With a number of degrees of freedom $p = 2$, specifying the (symmetric and positiv-semidefinite) covariance matrix requires $m = \frac{1}{2}p(p+1) = 3$ separate, independent entries, thus $t_{\mathbf{S}}$ has a χ^2_m distribution if H_0 is true.

- Choosing the significance level α.
  Conventionally one would require a significance level α of five percent. Thus, if we would observe $t_{\bar{\mathbf{x}}} > \chi^2_{2,\alpha}$ or $t_{\mathbf{S}} > \chi^2_{3,\alpha}$ we could disprove the respective null hypothesis with an error-probability of at most α. Simply setting $\alpha = 5\%$ and using the corresponding values χ^2_α turns out to be too restrictive. Although the CIE$(L^*a^*b^*)$ color space used in conjunction with our robust estimation scheme results in stable estimates for region attributes over different views, still due to different illumination conditions, different surface orientations relative to the camera position and due to varying scanning results, regions belonging to corresponding line segments in different views do not show exactly the same color distribution in a^*-b^*-color space. Instead of using a hard coded value for α we therefore provide an automatic initialization step: our algorithm randomly picks a small number (~ 10) of line segments in one view. For these line segments all geometrically possible match partners in the other view are determined. For the resulting combinations we compute the quantities $t_{\bar{\mathbf{x}}}$ and $t_{\mathbf{S}}$ as given in equations (5) and (6) respectively. We now have to discard obviously wrong matches, stemming from chromatically incompatible but geometrically possible combinations. A match is discarded if its significance level in the corresponding χ^2-test would fall below 0.001. From the remaining values we determine the median values $\hat{t_{\bar{\mathbf{x}}}}$ and $\hat{t_{\mathbf{S}}}$ as estimates for the test values on the left hand side of equations (5) and (6).

By requiring chromatic similarity on at least on side, the set of geometrically possible match candidates $\{\mathbf{l}'\}$ gets reduced substantially. For all geometrically and chromatically possible pairs we now compute the 3D reconstruction of the (finite) 3D line segment, stemming from the common part of both involved 2D line segments. In the next section, we will investigate these remaining candidates in more detail.

3.4 Restricting the initial match hypotheses

A line segment $\mathbf{l}'$ in the second view can only be the line segment corresponding to a given line segment $\mathbf{l}$ in the first view provided $\mathbf{l}'$ and $\mathbf{l}$ have similar chromatic properties at least on one side. The chromatic similarity is only demanded for one (corresponding) side of the flanking regions, since the line segment might be an occluding edge for one of its neighboring 3D patches. We now investigate this initial set of match candidates in more detail by computing the normalized cross-correlation in the a^*-b^*-bands. In Figure (4, a) a possible configuration of two line segments $(\mathbf{l}, \mathbf{l}')$ and their corresponding 3D line reconstruction is depicted. The epipolar relation generates a point-wise one-to-one correspondence between the corresponding lines $\mathbf{l}$ and $\mathbf{l}'$ in the two views; provided $\mathbf{l}'$ is not an epipolar line in the second view. Explicitly, a point $\mathbf{p}$ on line $\mathbf{l}$ corresponds to the point $\mathbf{p}'$ on $\mathbf{l}'$ in the second image, which is the intersection of $\mathbf{l}'$ with the epipolar line $\mathbf{l}'_{e,\mathbf{p}}$ of $\mathbf{p}$ in the second image.

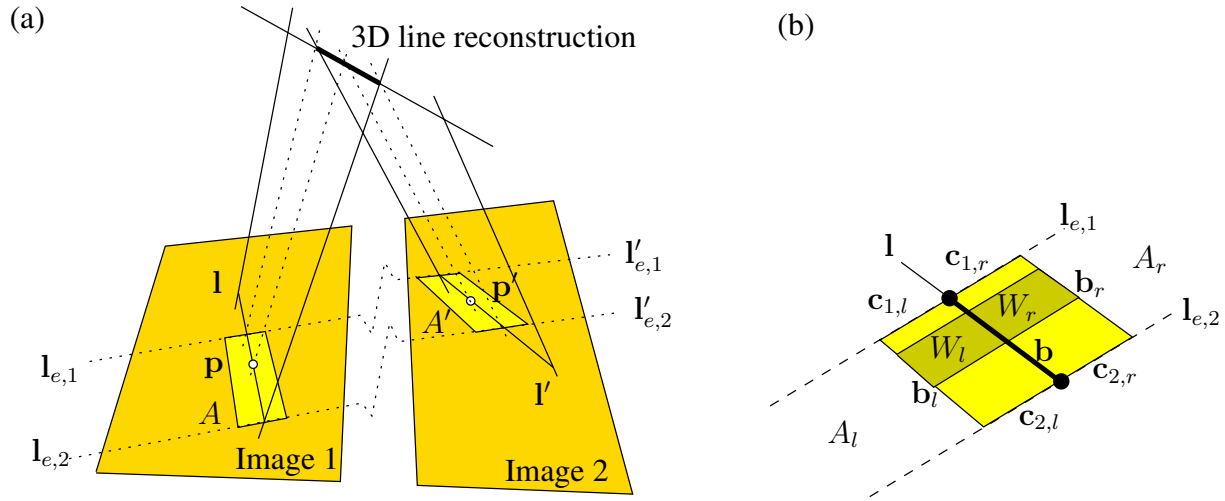


Figure 4: (a): Possible configuration when computing the 3D information for a candidate pair $(\mathbf{l}, \mathbf{l}')$. To compute the cross-correlation measure, the extent of the reconstructed 3D line segment is chosen to be the common part of the back-projections of $\mathbf{l}$ and $\mathbf{l}'$. Note the length of the common pieces in the two views can differ, due to perspective foreshortening. (b): Detail of the two-sided image area $A = \{A_l, A_r\}$ used for correlation and the two-sided correlation window $W = \{W_l, W_r\}$.

To be meaningful, the cross-correlation has to be computed over corresponding image regions. For each line segment pair $(\mathbf{l}, \mathbf{l}')$ forming a possible match, we therefore first determine the corresponding (two-sided) image areas A, A' in each view. For example, the outline of the right image area A_r in image one, belonging to line segment $\mathbf{l}$ is given by: the backprojection $\mathbf{b}$ of the common part of the 3D line reconstruction onto $\mathbf{l}$, the line segment $\mathbf{b}_r$ parallel to $\mathbf{b}$, and the two line segments $\mathbf{c}_{1,r}$ and $\mathbf{c}_{2,r}$ on the epipolar lines through the endpoints of $\mathbf{b}$.

Choosing the orientation of the sides $\mathbf{c}_{1,l/r}$ and $\mathbf{c}_{2,l/r}$ to be given by the epipolar lines turns out to be a reasonable approximation for restricting the image areas in the different views to one corresponding roof patch in the 3D scene. See below for cases, where the angle between the line segment and the epipolar lines from the second view is too small. If we further assume that the epipolar lines determining $\mathbf{c}_{1,l/r}$ and $\mathbf{c}_{2,l/r}$ are parallel to each other, which is a good approximation in the given framework, then the extraced image areas A, A' become parallelograms.

The length $c_{l/r} = |\mathbf{c}_{1,l/r}| = |\mathbf{c}_{2,l/r}|$ is variable. During the examination of a putative pair, the length $c_{l/r}$ of one of the two image areas A, A' is changed over a given interval to handle possible foreshortenings occurring to planar surfaces under projective perspective.

We obtain the pixels in the image areas A, A' by first transforming the left and right neighborhoods of $\mathbf{l}$ and $\mathbf{l}'$ in an axis aligned coordinate system, the same way as described in Section (3.2.1). Again, we ensure the area is not overlapping neighboring contours. During the affine transformation we additionally introduce a length scaling, taking into account the different length of the common pieces $\mathbf{b}$ and $\mathbf{b}'$ of the projected 3D line segment in the two images. The scaling is adjusted to map the transformed regions onto rectangles of the same length. Thus we make the implicit one-to-one correspondence induced from epipolar geometry directly accessible.

Let us now define the *pointwise* chromatic similarity measure h. The pointwise chromatic similarity measure h is the sum of the normalized cross-correlation in the individual bands a^* and b^*, evaluated at the point $\mathbf{p} = (x, y)$ in the first view and its corresponding point $\mathbf{p}' = (x', y')$ in the second view

$$h = \mathrm{ncc}_a(\mathbf{p}, \mathbf{p}') + \mathrm{ncc}_b(\mathbf{p}, \mathbf{p}')\,; \tag{7}$$

with $\mathrm{ncc}_k(\mathbf{p}, \mathbf{p}')$ being the normalized cross-correlation

$$\mathrm{ncc}_k(\mathbf{p}, \mathbf{p}') = \frac{\sum_{(i,j)\in W}\left[k(x+i, y+i) - \bar{k}(x, y)\right]\left[k'(x'+i, y'+i) - \bar{k'}(x', y')\right]}{\sqrt{\sum_{(i,j)\in W}\left[k(x+i, y+i) - \bar{k}(x, y)\right]^2}\sqrt{\sum_{(i,j)\in W}\left[k'(x'+i, y'+i) - \bar{k'}(x', y')\right]^2}} \tag{8}$$

of the intensities in color band $k \in \{a^*, b^*\}$ for a given correlation window $W \in \{W_l, W_r\}$, as depicted in Figure (4, b). The pointwise similarity measure h is obtained by correlating the left and right sides of the line segments separately, taking into account only the side(s) with satisfy the statistical chromatic similarity constraints. More precisely, the pixels actually included in the correlation-mask are selected the same way as described in Section (3.2.1): the included pixels only approach neighboring contours up to a specified offset and L^*-outlier pixels are not considered.

For the further reduction of match candidates, we employ the average $\bar{h}$ of the pointwise similarity measure h along the common part of both line segments. A pair is kept, if $\bar{h}$ is positive on the side(s) which satisfied the statistical chromatic similarity constraints.

Until now, we restricted ourselves to use only two input images. The benefits of using three ore more views can be seen in the results presented in (Moons et al., 1998, Baillard et al., 1999). One advantage in using three views is the availability of a strong geometric constraint: the trifocal tensor enables the prediction of a line segment reconstructed from two views in the third view. In this paper, we don't exploit the trifocal constraint; the third view only enters in the computation of the cross-correlation measure. If the third view is selected from a parallel flight line, as sketched in Figure (5), we can already overcome the limitations in delimiting the image areas used for cross-correlation. As mentioned above, the orientation of the sides $\mathbf{c}_{1,l/r}$ and $\mathbf{c}_{2,l/r}$ of the image area $A_{l/r}$ in Figure (4, b) is given by the epipolar lines through the common part $\mathbf{b}$ of $\mathbf{l}$. For the cases, where the angle between the line segment in image one and the epipolar line from image two is less than 40 degrees, we match this line segment against the line segments from image three.

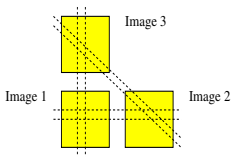


Figure 5: Schematic representation of the given input image configuration when considering a third view. The orientation of the epipolar lines is sketched with dashed lines.

4 RESULTS AND CONCLUSION

We tested our algorithm on a state-of-the-art-dataset, produced by Eurosense Belfotop n.v. It consists of high resolution color images of densely built up urban areas in Brussels. The image characteristics are: 1:4000 image scale and geometrically accurate film scanning with 20 microns pixel size (corresponding $8 \times 8\ cm^2$ on ground) , four-way image overlap, and precise sensor orientation. For visualization purposes the results obtained by the succeeding pruning steps are depicted in Figure (6) only for one building. In case (6, c, bottom row), the matched lines stemming from image two and image three were merged together in the right hand side picture.

The algorithm was applied to cut-outs with a size of approximately 500×500 pixels. The number of extracted line segments with a minimum length of 15 pixels, using the Maximum method described above was about 350 in each view. If we only restrict the maximal angle difference and the maximal allowed disparity to reasonable values, the number of all geometrically possible pairs is about 3700. This number could be reduced to nearly 1000, using only the the integrated chromatic properties of the flanking regions. If we additionally apply the chromatic cross-correlation constraint, we achieve a further reduction to approximately 700 candidates, also *without* loosing a single correct candidate.

The results demonstrate the efficient and reliable complexity reduction achieved by exploiting color information for line segment matching. Future work will focus on integrating a third view – and if available further views – seamlessly into the presented framework. We expect an even stronger discriminative power of the method by bringing together trifocal constraints and the developed chromatic similarity measures. The matched line segments, in conjunction with the extracted chromatic neighboring information will form the starting point for the following reconstruction steps towards the automatic generation of 3D city models.

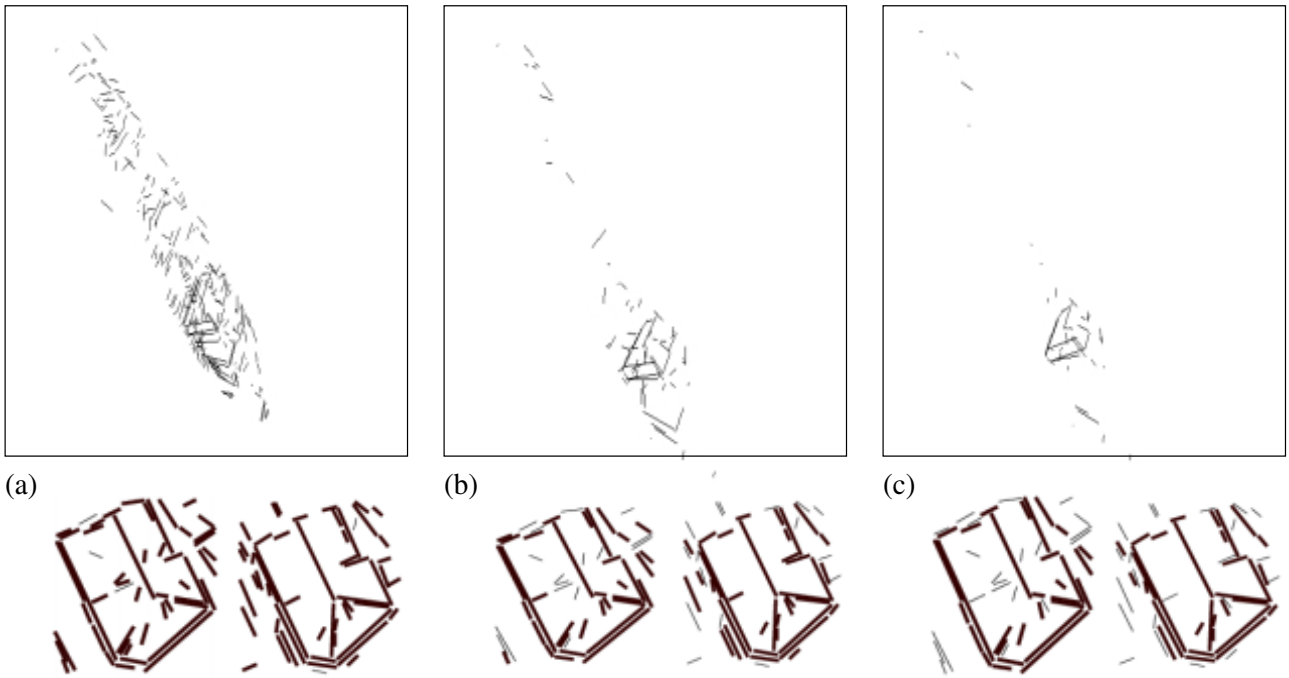


Figure 6: Top row: 3D reconstruction results for one building. Bottom row, bold: line segments in subimages one and two, that were matched (possibly multiply) over the views. (a): all geometrically possible matches (266 matches). (b): set of matches after investigating chromatic region properties (73 matches). (c): Remaining matches after additionally applying the cross-correlation constraint (42 matches).

REFERENCES

Baillard, C., Schmid, C., Zisserman, A. and Fitzgibbon, A., 1999. Automatic Line Matching and 3D Reconstruction of Buildings from Multiple Views. In: ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS Vol.32, Part 3-2W5, pp. 69–80. 7

Canny, J., 1986. A Computational Approach to Edge Detection. In: PAMI, pp. 679–698. 2

Deriche, R., 1987. Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector. In: International Journal of Computer Vision, pp. 167–187. 2

Guttman, A., 1981. R-Trees, a dynamic index structure for spatial searching. Proceedings of the SIGMOD Conference pp. 47–57. 2

Henricsson, O., 1996. Analysis of Images using Color Attributes and Similarity Relations. PhD thesis, No. 11663, Communication Technology Laboratory, Computer Vision Group, ETH Zürich, Switzerland. 5

Henricsson, O., 1998. The Role of Color Attributes and Similarity Grouping in 3-D Building Reconstruction. Computer Vision and Image Understanding 72(2), pp. 163–184. 4

Krzanowski, W. J., 1993. Principles of Multivariate Analysis. Oxford University Press. 5

Lee, H.-C., 1991. Detecting Boundaries in a Vector Field. IEEE Transactions in Signal Processing 39(5), pp. 1191–1194. 2

Moons, T., Frere, D., Vandekerckhove, J. and Gool, L. V., 1998. Automatic Modelling and 3D Reconstruction of Urban House Roofs from High Resolution Aerial Imagery. ECCV 1, pp. 410–425. 7

Rousseeuw, P., 1985. Multivariate Estimation with High Breakdown Point. In: W. Grossmann, G. Pflug, I. Vincze and W. Wertz (eds), Mathematical Statistics and Applications, Reidel Publishing Company (Dordrecht), pp. 283–297. 5

Rousseeuw, P. and Leroy, A. M., 1987. Robust Regression and Outlier Detection. John Wiley & Sons. 5

Schmid, C. and Zisserman, A., 1997. Automatic Line Matching Across Views. In: Proc. CVPR, pp. 666–671. 4

Zafiropoulos, P. and Schenk, T., 1999. Color-Based Contour Strategies for Road Extraction. In: IAPRS, Vol. 32. 2