Reprint from VISAPP 2007 Second International Conference on Computer Vision Theory and Applications, pp. 264-267, Barcelona, March 2007.

OBJECT RECOGNITION AND POSE ESTIMATION ACROSS ILLUMINATION CHANGES

Muselet D.
Laboratoire LIGIV EA 3070 - Université Jean Monnet - France
damien.muselet@univ-st-etienne.fr

Funt B., Shi L.
School of Computing Science, Simon Fraser University, Vancouver, Canada
funt@cs.sfu.ca, lshia@cs.sfu.ca

Macaire L.
Laboratoire LAGIS UMR CNRS 8146 - Université des Sciences et Technologies de Lille - France
ludovic.macaire@univ-lille1.fr

Keywords: Color histograms, object recognition, 2D pose estimation, illumination changes, local color descriptors

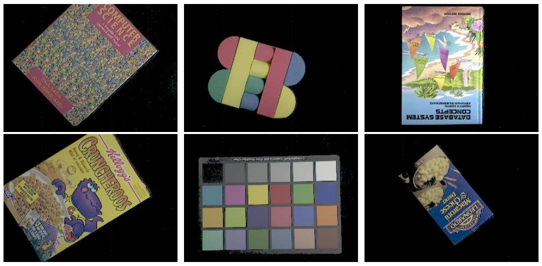
Abstract: In this paper, we present a new algorithm for color-based object recognition that detects objects and estimates their pose (position and orientation) in cluttered scenes observed under uncontrolled illumination conditions. As with so many other color-based object-recognition algorithms, color histograms are also fundamental to our approach; however, we use histograms obtained from overlapping subwindows, rather than the entire image. Furthermore, each local histogram is normalized using greyworld normalization in order to be as less sensitive to illumination as possible. An object from a database of prototype objects is identified and located in an input image by matching the subwindow contents. The prototype is detected in the input whenever many good histogram matches are found between the subwindows of the input image and those of the prototype. In essence, normalized color histograms of subwindows are the local features being matched. Once an object has been recognized, its 2D pose is found by approximating the geometrical transformation most consistently mapping the locations of prototype's subwindows to their matched subwindow locations in the input image.

1 INTRODUCTION

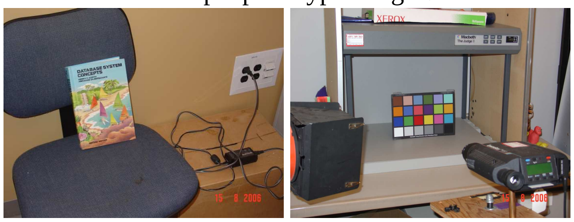
Starting with Swain and Ballard's color indexing (Swain and Ballard, 1991), color has proved to be a very important clue for object recognition. Following in this tradition, we present a new algorithm for color-based object recognition that detects objects and estimates their pose (position and orientation) in cluttered scenes under uncontrolled illumination conditions. As with so many other color-based object-recognition algorithms (Funt and Finlayson, 1995; Bressan et al., 2003), color histograms are also fundamental to our approach; however, we use histograms obtained from overlapping subwindows, rather than the entire image. Furthermore, each local histogram is normalized using greyworld normalization (Buchsbaum, 1980).An object from a database of prototype objects is identified and located in an input image by matching the subwindow contents. The prototype is detected in the input whenever many good histogram matches are found between the subwindows of the input image and those of the prototype. In essence, normalized color histograms of subwindows are the local features being matched. Once an object has been recognized, its 2D pose is found by approximating the geometrical transformation most consistently mapping the locations of prototype's subwindows to their matching subwindow locations in the input image (Lowe, 1999).

An entry in the database of prototypes is built from an image of a single object placed on an uniform background. The test images containing the objects to be recognized may contain several objects, some of which may be partially occluded. The prototype and test images are acquired under different illumination conditions and with the same zoom parameters (See Fig.1).

Color histograms are very effective for object recognition (Swain and Ballard, 1991) and image indexing (Park et al., 1999) because they are simple and fast to compute, are invariant to rotation and translation, and are insensitive to partial object occlusion. However, color histograms of whole objects are useless for determining object pose, precisely because they are rotation invariant. In terms of accurately determining object position, Swain's histogram back-



Example prototype images.



Example input test images.

Figure 1: These images come from the Simon Fraser University Database (Barnard et al., 2002) available at http://www.cs.sfu.ca/∼colour/data.

projection (Swain and Ballard, 1991) is very sensitive to noise and provide only a coarse estimate of the object's position. When the input images may contain several objects and include the possibility of partial occlusion, many other non-color-based approaches (Lowe, 1999; Ohba and Ikeuchi, 1997) rely on local image features for matching. The approach proposed here combines the ideas of color histogram matching and local feature matching.

Many of the local-feature-based object recognition methods (Bressan et al., 2003; Lowe, 1999; Ohba and Ikeuchi, 1997) extract interest points from the images as an initial step and then evaluate local descriptors around these points. A drawback of this approach is that the robustness and the repeatability of the interest-point detector becomes crucial. To avoid the reliability problems associated with interest-point detectors, we propose to analyze all the local neighborhoods in the image and to extract descriptors for all of them. Indeed, rather than to extract features from a limited number of areas in the image, we consider all pixels to be interest points and consequently extract the features around all of them. Thus we divide the image into overlapping subwindows and compute their color histograms. The computation in this step can be organized so that each pixel needs to be visited only once, so it is fast.

When the illumination is not controlled during the acquisition of the images, the classical color histograms lead to poor recognition results (Funt et al., 1998). Thus, we propose to normalize the color histograms in order to cope with this problem. One classical and computationally simple approach is greyworld normalization. The main drawback of this normalization is that it assumes that the illumination is spatially constant over the whole image (Buchsbaum, 1980). In normalizing each subwindow separately, we only assume that the illumination is constant within each subwindow, not across the whole image.

The database of prototypes represents each object in terms of the normalized local color (NLC) histograms from its image's subwindows. This representation associates a point on the object with each NLC histogram. To identify the objects in the input test image, each test-image subwindow is matched to the entire set of database subwindows and labeled according to the one that matches the best. A subset of subwindows with the same object label indicates the presence of the corresponding object in the image. The locations of the subwindow labels within the image indicates the object's pose.

The second section of this paper is about the illumination changes and the greyworld normalization. The third section presents details about how the space and time requirements for NLC histogram storage and matching can be reduced using incremental principal components analysis (Hall et al., 1999). The specifics of 2D-pose estimation are described in the fourth section. Results of tests based on the Amsterdam image database (Geusebroek et al., 2005) are given in the fifth section.

2 ILLUMINATION CHANGES

In order to deal with variations in the spectral composition of the incident illumination, we adopt the diagonal model of illumination change. The diagonal model assumes that the spectral sensitivity function of each sensor of the camera is sufficiently narrowband that they can be viewed as Dirac delta functions at three distinct wavelengths. In practice, although this assumption does not hold perfectly it is generally an adequate model, and it can be improved by spectral sharpening (Finlayson et al., 1994).

Using the diagonal model of illumination change along with the additional assumption that all pixels in a subwindow are lit by the same illumination, we can apply the greyworld normalization to each local color histogram by dividing each color component by its mean value within this subwindow. Each subwindow within then is characterized by a normalized local color (NLC) histogram.

3 EIGEN NORMALIZED LOCAL COLOR HISTOGRAMS

Since each prototype image represents only one object, each subwindow represents a specific area of the object. Thus, considering that we have P prototype images I^i_{pro}, $i \in \{1,...,P\}$, the prototype image I^i_{pro} which represents the object O^i, is divided into WP^i subwindows wp^i_j, $j \in \{1,...,WP^i\}$, each one representing the j^{th} area Op^i_j of the object O^i.

Since the proposed object recognition method requires the storage and matching of many subwindow histograms, it is important to reduce the memory and computation requirements as much as possible. One strategy for decreasing the complexity of histogram matching is to reduce the dimensionality of the histograms (Bressan et al., 2003; Tran and Lenz, 2005). Therefore, we apply principal component analysis to the set of prototype local color histograms.

Following the method of Tran and Lenz (Tran and Lenz, 2005), PCA is applied to histogram differences, rather than the histograms themselves. The histogram differences suffice since the aim when compressing histograms for the object recognition is not to be able to reconstruct the histograms, but only to estimate distances between histograms. Therefore, PCA applied on the space of histogram differences should lead to better results than PCA applied on the histogram space. Since we care only about similar images, PCA is not applied on all the differences between the prototype histograms, but only on the differences between similar prototype histograms. Hence, for each prototype histogram, we use the histogram difference between it and its closest prototype histogram from the same image. The closest histogram is the one at the minimum Manhattan distance between the two histograms. Swain showed that the Manhattan distance is equivalent to use of the intersection between color histograms when these histograms contain the same number of pixels (Swain and Ballard, 1991).

When the number of images in the prototype database is high, the number of NLC histograms becomes very high and the time required to apply principal component analysis becomes prohibitive. To overcome this limitation, we move to incremental PCA (Hall et al., 1999). Thanks to incremental PCA, the size of the prototype database is effectively unlimited. The IPCA step is completed once off-line.

The NLC histograms projected onto the eigenbasis from IPCA are then called eigen NLC (ENLC) histograms. All NLC histograms, from both the database of prototypes and the input test image, are projected onto the same eigenbasis. Finally, each input ENLC histogram is compared against all the prototype ENLC histograms, and the most similar prototype ENLC histogram is kept. Histograms are matched according to the Manhattan distance between them.

After the matching step, each sub-window wq_k, $k \in \{1,...,WQ\}$, of the input image is associated with one prototype subwindow wp^i_j, and so, with one object area Op^i_j of the object O^i. The subwindow's labels $[input\ subwindow,\ object\ area]$ $[wq_k, Op^i_j]$ are used to determine the best geometrical transformation mapping the corresponding prototype image to the input image.

4 2D POSE ESTIMATION

After the matching step, the subwindows wq_k, $k \in \{1,...,WQ\}$, from the input image will have an associated object area Op^i_j. Let C^i, $i \in \{1,...,T\}$, $T \leq WP^i$, denote the subset of features (areas) from the object O^i that have been associated with at least one input subwindow wq_k: $C^i = \{Op^i_j | there\ exists\ k\ so\ that\ [wq_k, Op^i_j]\ exists\}$. We next consider the non-empty subsets C^i one by one and estimate the orientation and position of the corresponding object O^i in the input image. This means finding the geometric transformation from the spatial coordinates of the object O^i in the prototype image I^i_{pro} to its coordinates in the input image. The estimation of this transformation is based on the spatial coordinates of the matching subwindow pairs.

As described by Lowe (Lowe, 1999), the geometric transformation from a point $[x,y]^T$ associated with a prototype subwindow to a point $[u,v]^T$ associated with the corresponding input subwindow can be written as:

$$\begin{bmatrix} x & -y & 1 & 0 \\ y & x & 0 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ t_x \\ y_y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} \quad (1)$$

where t_x and t_y represent the translation parameters and the m_i represent the rotation around the center of the object and scale parameters.

This equation is based on one pair of prototype and input subwindows, but we can add some other pairs and calculate the least-squares solution for the geometric parameters.

Since this method is very sensitive to outliers, we propose the following two step approach:

- The set C^i is randomly divided into subsets of a fixed number of features, and then the least-squares fit for the geometric transformation for each of these subsets is determined independently. A large residual error in the fit indicates mismatched features that are then deleted from the set C^i. The number of features in a subset and the threshold of the residual error are fixed parameters determined experimentally.

- Using only those features leading to low residual error in the preceding step, the best geometric transformation is determined by least-squares fitting.

5 EXPERIMENTAL RESULTS

We first test our algorithm on the real images from Fig.1 and the recognition and pose estimation are perfect. Then, the Amsterdam Library of Object Images (ALOI) database (Geusebroek et al., 2005) is used for testing. The Amsterdam database contains 12 sets of color images. Each set contains images of one object on a uniform background under one of the 12 different illuminants having color temperatures between 2175°K to 3075°K. For the tests, we use 2 sets of color temperature 2325°K and 2750°K. 250 images of the first set are used as the prototype images. From the second set, we extract 100 objects to create 20 input images, each one representing 5 objects. Each object is subject to 2D rotation and translation before being added to the set of input images.

For these tests, the size of the subwindows is fixed at 45x45 pixels, and the offset between the centers of two neighboring subwindows is 15 pixels. The average number of ENLC histograms for each prototype image is 250. The number of bins in a raw histogram is $8^3 = 512$. After projection on the eigenbasis, this number reduces to 64.

The algorithm correctly recognizes and makes a perfect estimate of the pose for 96 of the 100 input objects.

6 CONCLUSION

A method for object recognition and 2D pose estimation has been presented. The method is insensitive to the color of the scene illumination. The basic strategy is to match local image features, in particular, to match the color histograms of subwindows from the input image to histograms of subwindows of prototypes in the database. The subwindow contents are normalized via greyworld averaging to remove the effects of variations in illumination. Pose is determined by finding the best correspondences between the matching subwindows that are consistent with a single geometrical transformation. Overall the accuracy of the proposed method is quite good considering that the database comprises images of objects with quite similar color distributions imaged under lights of different color temperature than the input images.

REFERENCES

Barnard, K., Martin, L., Funt, B., and Coath, A. (2002). A data set for colour research. *Color Research and Application*, 27(3):147–151.

Bressan, M., Guillamet, D., and Vitria, J. (2003). Using an ica representation of local color histograms for object recognition. *Pattern Recognition Letters*, 36:691–701.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Jour. of the Franklin Institute*, 310:1–26.

Finlayson, G., Drew, M., and Funt, B. (1994). Color constancy: Generalized diagonal transforms suffice. *Jour. of the Optical Society of America A*, 11(11):3011–3020.

Funt, B., Barnard, K., and Martin, L. (1998). Is machine colour constancy good enough? In *Procs.of the* 5^{th} *European Conf. on Computer Vision*, pages 445–459.

Funt, B. and Finlayson, G. (1995). Color constant color indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(5):522–529.

Geusebroek, J. M., Burghouts, G. J., and Smeulders, A. W. M. (2005). The amsterdam library of object images. *Int. Jour. of Computer Vision*, 61(1):103–112.

Hall, P., Marshall, D., and Martin, R. (1999). Adding and subtracting eigenspaces. *British Machine Vision Conference*, 2:463–472.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Procs. of the International Conf. on Computer Vision*, pages 1150–1157, Corfou.

Ohba, K. and Ikeuchi, K. (1997). Detectability, uniqueness, and reliability of eigen windows for stable verification of partially occluded object. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19 (9):1043–1048.

Park, D., Park, J., Kim, T., and Han, J. (1999). Image indexing using weighted color histogram. In *Procs. of ICIAP*, pages 909–914.

Swain, M. J. and Ballard, D. H. (1991). Color indexing. *Int. Jour. of Computer Vision*, 7(1):11–32.

Tran, L. V. and Lenz, R. (2005). Compact colour descriptors for colour-based image retrieval. *Signal Process.*, 85(2):233–246.